Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 15, 2004, 22:25:40 ; Search time 12355 Seconds
(without alignments)
6083.619 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQTWRATEPRYP.........WDEEFKPLLCSQYETLSDSE 2517
Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop   6.0 , Fgapext 7.0
                Delop   6.0 , Delext   7.0
Searched: 27513289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09522753/runat_15042004_143737_17439/app_query.fasta_1.2695
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=100
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=100 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09522753 @CGN_1_1_10059 @runat_15042004_143737_17439 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
         1:     em_estba:*
         2:     em_esthum:*
         3:     em_estin:*
         4:     em_estmu:*
         5:     em_estov:*
         6:     em_estpl:*
         7:     em_estro:*
         8:     em_htc:*
         9:     gb_est1:*
        10:     gb_est2:*
        11:     gb_htc:*
        12:     gb_est3:*
        13:     gb_est4:*
        14:     gb_est5:*
        15:     em_estfun:*
        16:     em_estom:*
        17:     em_gss_hum:*
        18:     em_gss_inv:*
        19:     em_gss_pln:*
        20:     em_gss_vrt:*
        21:     em_gss_fun:*
        22:     em_gss_mam:*
        23:     em_gss_mus:*
        24:     em_gss_pro:*
        25:     em_gss_rod:*
        26:     em_gss_phg:*
        27:     em_gss_vrl:*
        28:     gb_gss1:*
        29:     gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8226	62.2	7372	29	AY412686	AY412686 Homo sapi
2	7109.5	53.8	7013	29	AY412688	AY412688 Mus muscu
3	6533	49.4	6836	29	AY412687	AY412687 Pan trogl
4	2566	19.4	2000	11	BC033087	BC033087 Homo sapi
5	2264	17.1	1362	11	BC020427	BC020427 Homo sapi
6	1643.5	12.4	1966	11	BC026028	BC026028 Homo sapi
7	1634	12.4	1850	11	BC035748	BC035748 Homo sapi
8	1626.5	12.3	1835	11	BC029627	BC029627 Mus muscu
9	1624.5	12.3	1819	11	BC062162	BC062162 Mus muscu
10	1624	12.3	1093	12	BM909096	BM909096 AGENCOURT
11	1603	12.1	1996	11	BC026623	BC026623 Mus muscu
12	1559	11.8	965	12	BM555371	BM555371 AGENCOURT
13	1540	11.7	1126	12	BM553310	BM553310 AGENCOURT
14	1480.5	11.2	1057	12	BM477568	BM477568 AGENCOURT
15	1479.5	11.2	1051	12	BM423558	BM423558 AGENCOURT
16	1471.5	11.1	937	13	BQ949384	BQ949384 AGENCOURT
17	1448	11.0	846	13	BU557144	BU557144 AGENCOURT
18	1442.5	10.9	984	13	BQ690869	BQ690869 AGENCOURT
19	1440.5	10.9	1159	12	BM802749	BM802749 AGENCOURT
20	1420	10.7	1022	12	BM910785	BM910785 AGENCOURT
21	1382.5	10.5	1030	12	BM558844	BM558844 AGENCOURT
22	1378.5	10.4	1010	12	BM915686	BM915686 AGENCOURT
23	1374	10.4	887	13	BU542258	BU542258 AGENCOURT
24	1368	10.4	1067	12	BM472005	BM472005 AGENCOURT
25	1366.5	10.3	868	13	BX368971	BX368971
26	1348.5	10.2	923	13	BU184403	BU184403 AGENCOURT
27	1346.5	10.2	1038	12	BM910704	BM910704 AGENCOURT
28	1338	10.1	887	14	CA979881	CA979881 AGENCOURT
29	1334	10.1	926	13	BX390462	BX390462
30	1332	10.1	875	13	BQ711119	BQ711119 AGENCOURT
31	1331.5	10.1	842	13	BU224569	BU224569 AGENCOURT
32	1326	10.0	1094	12	BM560912	BM560912 AGENCOURT
33	1321.5	10.0	905	13	BQ897825	BQ897825 AGENCOURT
34	1321	10.0	882	13	BU180236	BU180236 AGENCOURT
35	1312.5	9.9	943	13	BX368972	BX368972
36	1305.5	9.9	898	13	BU172348	BU172348 AGENCOURT
37	1305	9.9	1045	12	BM560255	BM560255 AGENCOURT
38	1274	9.6	875	13	BQ691710	BQ691710 AGENCOURT
39	1262.5	9.5	915	14	CF272442	CF272442 AGENCOURT
40	1260	9.5	1075	12	BM461469	BM461469 AGENCOURT
41	1256.5	9.5	928	13	BU164114	BU164114 AGENCOURT
42	1256	9.5	994	13	BX397973	BX397973
43	1253.5	9.5	1066	12	BM471347	BM471347 AGENCOURT
44	1252.5	9.5	797	14	CB723733	CB723733 UI-M-FY0-
45	1250	9.5	1087	12	BG252257	BG252257 AGENCOURT
46	1243	9.4	958	12	BG831424	BG831424 AGENCOURT
47	1239	9.4	810	12	BG974253	BG974253 AGENCOURT
48	1235	9.3	962	13	BQ944732	BQ944732 AGENCOURT
49	1231.5	9.3	1113	12	BM811122	BM811122 AGENCOURT
50	1227.5	9.3	861	13	BU191146	BU191146 AGENCOURT
51	1225.5	9.3	876	12	BG252161	BG252161 AGENCOURT
52	1224.5	9.3	763	14	CB248833	CB248833 UI-M-EX0-
c 53	1216	9.2	845	9	AL884718	AL884718 AL884718
54	1204	9.1	900	13	BQ214358	BQ214358 AGENCOURT
55	1195.5	9.0	1132	12	BM910968	BM910968 AGENCOURT
56	1183.5	9.0	893	13	BU186180	BU186180 AGENCOURT
57	1183	9.0	790	13	BU257626	BU257626 603414849
58	1175	8.9	1032	13	BQ070408	BQ070408 AGENCOURT
c 59	1161.5	8.8	858	13	BX732682	BX732682 BX732682
60	1158.5	8.8	787	10	BE793487	BE793487 601588814
61	1153.5	8.7	718	14	CA749602	CA749602 UI-M-FY0-
62	1151.5	8.7	704	14	CD355812	CD355812 UI-M-FY0-

```
FEATURES       Location/Qualifiers
source         1..7372
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
gene           <1..>7372
               /gene="NCOR2"
               /locus_tag="HCM4640"

ORIGIN

Alignment Scores:
Pred. No.:                      0                Length:        7372
Score:                          8226.00          Matches:       1708
Percent Similarity:             69.56%           Conservative:  1
Best Local Similarity:          69.52%           Mismatches:    748
Query Match:                    62.25%           Indels:        0
DB:                             29               Gaps:          0

US-09-522-753-5 (1-2517) x AY412686 (1-7372)

Qy    1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro   20
Db    1 ATGTCGGGATCCACACAGCTTGTGGCACAGACGTGGAGGGCCACCGAGCCCCGCTACCCG   60

Qy   21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu   40
Db   61 CCCCACAGCCCTTCCTACCCAGTGCAGATCGCCCGGACGCACACGGACGTCGGGCTCCTG  120

Qy   41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleGln   60
Db  121 GAGTACCAGCACCACTCCCGCGACTATGCCTCCCACCTGTCGCCCGGCTCCATCATCCAG  180

Qy   61 ProGlnArgArgProSerLeuLeuSerGluGlnPheGlnProGlyAsnGluArgSerGln   80
Db  181 CCCCAGCGGCGGCCGAGCCTCCTGCTCTGAGTTCCAGCCCGGGAATGAACGGNNNNNNN  240

Qy   81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProLeuGlyLeuLysSerGluMet  100
Db  241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  300

Qy  101 GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro  120
Db  301 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  360

Qy  121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuThrLysAspArgSer  140
Db  361 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACCGTAGC  420

Qy  141 LeuThrGlyLysGluGluProValSerProValSerProProHisThrAspProGluLeu  160
Db  421 CTGACGGGCAAGGTGGAACCGGTGTCTCCCCCAGCCCCCACACTGACCCTGAGCTG      480

Qy  161 GluLeuValProProArgLeuSerGluLysSerGluGluIleGlnAsnMetAspArgValAsp  180
Db  481 GAGCTGGTGCCGCCAAGGCTGTCCAAGGAGAGCTGATCCAGAACATGGACCGCGTGGAC  540

Qy  181 ArgGluIleThrMetValGlnGlnProGluLysSerLeuLysGlnGlnLeu  200
Db  541 CGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAGAAGCAGCAACAGCTG  600

Qy  201 GluGluGluAlaAlaAlaLysProArgGluLeuSerProValSerProProIleGlu  220
Db  601 GAGGAGGAGGCTGCCAAGCCGCCCGAGCCCTGTGTCACCGCCACCATCGAG  660

Qy  221 SerLysHisArgSerLeuValGlnIleIleGluTyrAspGluAsnArgLysLysAlaGluAla  240
Db  661 TCGAAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCCGNNNNNNNNNNNNNNNN  720

Qy  241 AlaHisArgIleLeuGluGlyLeuGlyTyrProGlnValGluLeuTyrAsnGlnPro  260
Db  721 NNNNNNNNNNNNNNNNNNNNNNNNNNNCGCTGTACAACCAGCCC  780

Qy  261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu  280
```

63 1124 8.5 804 10 BE728145 601563413
64 1123 8.5 779 12 BM944466 UI-M-EH0p
c   65 1116 8.4 811 13 BX771375 BX771375
66 1114 8.4 774 14 CB521565 UI-M-GH0-
67 1111 8.4 651 12 BM783748 K-EST0061
68 1110.5 8.4 683 14 CF540568 UI-M-GV0-
69 1107 8.4 622 9   AI830862 wj6lb09.x
c   70 1105 8.4 657 10 BF530324 602071630
71 1104.5 8.4 902 12 BG485444 602505328
c   72 1099.5 8.4 625 14 CB216442 NISC_nq05
73 1090 8.3 825 13 BX732681 BX732681
74 1090 8.3 922 10 BE542336 601067079
c   75 1087 8.2 682 13 BX869832 BX869832
76 1085 8.2 790 12 BI081252 602879279
77 1081 8.2 729 14 CA325207 UI-M-FY0-
78 1074 8.1 809 13 BU475754 603471578
79 1073.5 8.1 788 12 BM011980 603636389
80 1070.5 8.1 688 10 BE391091 601286040
c   81 1070 8.1 664 10 BF058781 7k34g07.x
82 1062.5 8.0 1088 13 BU108220 603112484
83 1053 8.0 592 10 BF436351 7p06e11.x
84 1039.5 7.9 820 12 BI735071 603356210
85 1038 7.9 611 10 AW438580 xs82e10.x
86 1038 7.9 624 12 BG391632 602417893
87 1038 7.8 668 14 CB518985 UI-M-GH0-
88 1037 7.8 925 14 CD244013 AGENCOURT
89 1035 7.8 1030 12 BG119261 602349134
90 1033.5 7.8 654 14 AW701437 uq86b05.y
91 1032 7.8 725 14 CA317171 UI-M-FW0-
92 1030 7.8 614 12 BM771666 K-EST0055
93 1027.5 7.8 810 10 BF606925 MY2_00011
c   94 1025 7.8 612 9   AI523558 th08c09.x
95 1020.5 7.7 898 10 BF339827 602038834
96 1020.5 7.7 637 12 BI428444 fr84a07.y
97 1017.5 7.7 760 13 BU143568 603229009
98 1014 7.7 612 10 AW674918 bb30f04.y
99 1012 7.7 610 10 BE675456 7f09d11.x
100 1011.5 7.7 676 13 BY763864 BY763864

ALIGNMENTS

RESULT 1
AY412686
LOCUS   AY412686   7372 bp   DNA   linear   GSS 16-DEC-2003
DEFINITION Homo sapiens NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence.
            genomic survey sequence.
ACCESSION   AY412686
VERSION   AY412686.1 GI:39768651
KEYWORDS   GSS.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 7372)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL   Science 302 (5652), 1960-1963 (2003)
PUBMED   14671302
REFERENCE   2 (bases 1 to 7372)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE   Direct Submission
JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment.

```
Db  1861 GAAACAGCCAAGAAAGGTCTCCTGGAACACGGCGCGCCAACTGGTCGGCCATCGCCCGGATG 1920

Qy   641 ValGlySerLysThrValSerGlnPheTyrPheThrTyrLysLysArgGln           660
         |||||||||||||||||||||||||||| |||:::||||||||||||||||
Db  1921 GTGGGCTCCAAGACTGTGTCGCAGTGTAAGAACTTCTACTTCAACTACAAGAAGAGGCAG 1980

Qy   661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLysMetGluLysGluArgAsnAlaArg   680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 AACCTCGATGAGATCTTGCAGCAGCACAAGCTGAAGATGGAGAAGGAGAGAAACGCGCGG 2040

Qy   681 ArgLysLysGlySerLysAlaProAlaAlaAlaSerGluGluAlaAlaPheProProValVal 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 AGGAAGAAGAAAGGCGCGCCAGCGGCCGCGGCCAGCGAGGAGGCTGCATTCCCCGCCGTGGTG 2100

Qy   701 GluAspGluMetGluLeuSerGlyValSerGlyLysValGlnArgGlyGlyMetValGluGlu 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 GAGGATGAGGAGATGGAGCTGTCGGGCGTGAGCGGGAAAGTGAGGAGGAGATGGTGGAGGAG 2160

Qy   721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyGluCysSerGlyProAla   740
         |||||
Db  2161 GCTGAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2220

Qy   741 ThrValAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaAlaLys  760
Db  2221 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2280

Qy   761 AspThrGlyGlnAsnGlyProLysProProAlaThrLeuGlyAlaAspGlyProProPro  780
Db  2281 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340

Qy   781 GlyProProThrProProArgArgThrSerArgAlaProIleGluProThrProAlaSer  800
Db  2341 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2400

Qy   801 GluAlaThrGlyAlaProThrProProAlaProProSerProSerAlaProProPro     820
Db  2401 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2460

Qy   821 ValValProLysGlyGluLysGlyGluGluThrAlaAlaAlaProValGluGluGly     840
Db  2461 NNNNACAGTGGCTGAGTGCGTCCTCTATTACTACCTGACTAAGAAGAATGAGAACTATAAG 2520

Qy   841 GluGluGlnLysProProAlaAlaGluLeuAlaValAspThrGlyLysValAlaGluGlu   860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 GAGGAGCAGAAGCCCCCGGCTGAGGAGCTGCAGTGGACACAGGAGGCCGAGGAG        2580

Qy   861 ProValLysSerGluCysThrGluGluAlaGluGlyProAlaLaGlyLysAspAla       880
         ||||||||||||||||||||||||||||||||||||||||||
Db  2581 CCCGTCAAGAGCGAGTGCACGGAGGAGGCCGAGGGCCCAAGGCAAGGACGCG          2640

Qy   881 GluAlaAlaGluAlaThrAlaThrAlaProThrAlaAlaAlaProValGluGlyGlySer   900
Db  2641 GAGGCCGCTGAGGCCACGGCCACGGCGCCCACCGAGGAGGCAGAAGGAGGGGCGGGAGC   2700

Qy   901 GlyArgAlaThrAlaLysSerGlyAlaProGlnAspSerSerAlaThr             920
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701 GGCAGGGCCACCGCAAGAGCGGCGCCCCAGGACAGCAGCGCACTCCAGTGCTACC      2760

Qy   921 CysSerAlaAspGluValAlaGlyGluGlyProAlaLaGlyLysAspLysAsnArgLeuSerPro 940
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761 TGCAGTGCAGACGAGGTGATGAGCCGCGCGCCGACAGGGAAGGCCGGTCTGTCCCCA    2820

Qy   941 ArgProSerLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnGln        960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821 AGGCCCAGCTCCTCACCCCGACTGGCGACCCCGGGCAATGCCTCACCCCAGAAGCCA    2880

Qy   961 LeuAspLeuLysGlnLeuLysGlnArgAlaAlaAlaIleProProGlnValThrLys     980
         |||||||||||||||||||||||||||||||||||||||
Db  2881 CTGGACCTGAAGCAGCTGAAGCAGCGCGCCGCCATCCCCCCCCCCAGGTCACCAAA      2940

Qy   981 ValHisGluProArgGluAspAlaAlaProThrLysProProProAlaProPro       1000
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  2941 GTCCATGAGCGCCCCCGGGAGGACGCAGCAGCTCCCACCAAGCCAGCTCCCCCAGCCCCACCG 3000
```

```
Db   781 TCCGCACACCGGCAGTATCATGAGAACAATCAAAATAAACCAGGCAGCGATGCGGAGGAAGCTA 840

Qy   281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln    300
         |||||||||||||||||||||||||||||:::||||||||||||||||
Db   841 ATCTTGTACTTCAAGAGGAGAATCACGGTCGGAAACAATGGGACCAGAAGTTCTGCCAG     900

Qy   301 ArgTyrAspGluMetGluAlaLeuGluLysValLysValGluArgIleGluAsnAsnPro    320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CGCTATGACCAGCTCATGGAGGCCCTGGAGAAGAAGGTGGAGCGCATCGAGAACAACCCC    960

Qy   321 ArgArgArgAlaLysGluSerLysLysValArgGluValTyrTyrGluLysGlnPheGluIle 340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 CGGCGGCGGGCCAAGGAGCAAGGTGCGCGAGCAGAGCCAGTACTACGAGAAGCAGTTCCCTGAGATC 1020

Qy   341 ArgLysArgLeuGlnLnArgMetGlnSerArgValGlyGlnArgGlySerGly         360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CGCAAGCAGCGCGCTGCAGGCGCATGCAGCAGCTGGGCCAGCCCAGGCCAGTGGG      1080

Qy   361 LeuSerMetSerAlaAlaAsgGluHisGluGluIleIleAspGlyLeuSer           380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 CTGTCCATGTCCGCCGCCGCCCGCGAGCACGAGGTGTCAGAGATCATCGATGGCCTCTCA  1140

Qy   381 GluGlnGluLeuGlyLysGlnMetArgGlnLeuAlaValIleProMetLeuTyr        400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GAGCAGGAGAAGGGCAAGCAGATGCGCCAGCTGGCCGTGATCCCCATGCTGTAC       1200

Qy   401 AspAlaAspGlnAsnGlyIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMet   420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 GACGCTGACCAGCAGCCCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG  1260

Qy   421 LysValValTyrLysAspArgArgGlnValMetAsnMetTrpSerGluGlnGluLysThrPhe 440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 AAGGTGTACAAAGACCGCCGACCAGTCATGAACATGTGGAGTGAGCAGGAGAAGACCTTC  1320

Qy   441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg   460
         ||||||
Db  1321 CGGGAGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1380

Qy   461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys   480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 NNNACAGTGGCTGAGTGCGTCCTCTATTACTACCTGACTAAGAAGAATGAGAACTATAAG 1440

Qy   481 SerLeuValArgArgSerTyrArgArgArgGlyLysSerGlnGlnGlnGlnNNNNNNNNNN 500
         |||||||||||||||||||||||||||||||||||
Db  1441 AGCCTGGTGAGACGGAGCTATCGCGGCCGCGGGAAGNNNNNNNNNNNNNNNNNNNNNNNN  1500

Qy   501 GlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerGlnGluGluLys        520
         |||||||||||||||||||||||||||||||||||NNNNNNNNNNNNNNNNNNN
Db  1501 GGGAGGAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560

Qy   521 AspGlyLysLysGluLysGluAlaGluGluGluLysProGluValGluAsn          540
         |||||||||||||||||||||||||||||||||||||||||NNNNNNNNNN
Db  1561 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620

Qy   541 AspLysGluAspLysLysGluLysLysThrAspAspSerGlyValLysGlyLysGlyGly   560
         ||||||||||||||||||||||||||||||||||||
Db  1621 NNNNNNNNNNNNNNNNNNNNGGAAGAGACAGAGCAGCCAAAACTGCCAACAACGCAGCAG 1680

Qy   561 LysGluAlaValAlaAlaSerLysLysGlyArgLysLysThrAlaAsnSerGlnAsnGlnGlyArgArgLysGly 580
         |||||||||||||||||||||||||||||||||||||||||||
Db  1681 AAGGAGGCTGTGCCCGCTCAATGCTAATGAGGCCAACACTGCCACCAGCCAGAAGACGCAAAGGC 1740

Qy   581 ArgIleThrArgSerMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnGln   600
         |||||||||||||||||||||||||||||||||||||||||
Db  1741 CGCATCACCCGCTCAATGGCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCCAGCAG   1800

Qy   601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet     620
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 AGCGCCGAGCTGGCCTCCATGGAGCTGAATGAGAGTTCTCGCTGGACAGAAGAAAATG   1860

Qy   621 GluThrArgAlaLysGluLeuLeuHisGlyArgAsnTrpSerArgAlaIleAlaArgMet   640
         ||||||||||||||||||||||||||||||||||||||||
```

```
Qy  1001  ProProGlnAsnLeuGlnProGluSerAspAlaProGlnProGlySerProArg  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  CCACCGCAAAACCTGCAGCCGGAGAGCGACGCCCCTGGCAGCAGCCCCCG  3060

Qy  1021  GlyLysSerArgSerProAlaProProAlaAspLysProAlaHisGluAlaAlaGluAlaGln  1040
          ||||||||||||||||||||||||||||||||||
Db  3061  GGCAAGAGCAGGAGCCCGGCACCCCCGCGACAAGGAGNNNNNNNNNNNNNNNN  3120

Qy  1041  LysLeuProGlyAspProCysTrpThrSerGlyTyrLeuProPheProValProProArg  1060
Db  3121  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  3180

Qy  1061  GluValIleLeuLysAlaSerProHisAlaProAspProSerAlaPheSerTyrAlaProPro  1080
Db  3181  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  3240

Qy  1081  GlyHisProLeuProLeuGlyTyrLeuHisAspThrAlaArgProValLeuProArgProPro  1100
          ||||||||||||||||||||||||||||||||||||||||||
Db  3241  NGTCACCCACTGCCTGGGCCTCCATGACACTGCCCGGCCGTCCTGCGCCGCCCCACCCC  3300

Qy  1101  ThrIleSerAsnProProProLeuIleSerSerAlaLysHisProSerValLeuGluArg  1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3301  ACCATCTCCAACCCGCCTCCCCTCATCTCCTCTGCCAAGCACCCCAGCGTCCTCGAGAGG  3360

Qy  1121  GlnIleGlyAlaIleSerGlnGlyMetSerValGlnHisValProTyrSerGluHis  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3361  CAAATAGGTGCCATCTCCCAAGGAATGTCGGTCCAGCATGTCCCGTACTACCAGAGCAT  3420

Qy  1141  AlaLysAlaProValGlyProValThrMetGlyLeuProLeuProMetAspProLysLys  1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3421  GCCAAGGCCCCGGTGGGCCCTGTCACCATGGGGCTGCCCCTGCCCATGGACCCCAAAAAG  3480

Qy  1161  LeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyPro  1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3481  CTGGCACCCTTCAGCGGAGTGAAGCAGGAGCAGCTGTCCCCACGGGGCCAGGCTGGGCCA  3540

Qy  1181  ProGluSerLeuGlyValProThrProValThrAlaGlnAlaSerValLeuArgGlyThrAlaLeu  1200
          ||||||||||||||||||||||||||||||||||||||||||||
Db  3541  CCGGAGAGCCTGGGGGTGCCCACAGCCGTCCGTGCTGAGAGGACAGCTCTG  3600

Qy  1201  GlySerValProGlyGlySerIleThrLysThrGlyIleProSerThrArgValProSerAsp  1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3601  GGCTCAGTTCCGGGCCGGAAGCATCACCAAAGGCATTCCCAGCACACGGGTGCCCTCGGAC  3660

Qy  1221  SerAlaIleThrTyrArgGlySerIleThrGlyIleAspSerProAlaAspValLeuTyrLys  1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3661  AGCGCCATCACATACCGCGGCTCCATCACCGGCACGCACGGCCAGCTCCTGTACAAG  3720

Qy  1241  GlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgArgGlyArgGlu  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3721  GGCACCATCACCAGGATCATCGGCGAGGACAGCCCCAGTCGCTTGGACCCGCGGGGCCGGGAG  3780

Qy  1261  AspSerLeuProLysGlyHisValIleGluTyrGluGlyLysLysGlyHisValLeuSerTyr  1280
          ||||||||||||||||||||||||||||||||||||||||||||
Db  3781  GACAGCCTGCCCAAGGGCCACGTCATCGAGTACGAAGGGCCACGTCTTGTCCTAT  3840

Qy  1281  GluGlyGlyMetSerValThrGlnCysSerSerLysGluAspGlyArgSerSerGlyPro  1300
          |||||
Db  3841  GAGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  3900

Qy  1301  ProHisGluThrAlaAlaProLysAspArgThrTyrAspMetMetGluGlyTyrGlyArgValGlyArgArg  1320
Db  3901  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  3960

Qy  1321  AlaIleSerSerAlaSerIleGluGlyValLeuMetGlyValArgAlaIleProGluProGluArgHis  1340
          ||||||||||||||||||||||||||||
Db  3961  NNNNNNNNNNNNNNNGTCTCATGGGTGCCATCCCGCTGCCGAGCGACAC  4020

Qy  1341  SerProHisHisLeuHisGluGlnHisHisIleHisIleArgGlySerIleThrGlnIleTyrLeuPro  1360
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  4021  AGCCCCCCACCTCAAAGACAGCAGCACCACCACCATCCACATCCGGGGGTCCATCACACAAGNNNNNNNN  4080

Qy  1361  ArgSerTyrValGlnAlaGlyGlnGluAspTyrLeuLysArgArgGluArgGluAlaAlaLysLeuLeuLysArg  1380
Db  4081  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4140

Qy  1381  GluGlyThrProProProProProSerArgAspLeuThrGluAlaTyrLysThrGln  1400
Db  4141  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4200

Qy  1401  AlaLeuGlyProLeuLysProLeuLysProAlaHisGluGlyLeuGlyLeuAlaAlaThrValLysGlu  1420
Db  4201  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4260

Qy  1421  AlaGlyArgSerIleHisGluIleProArgGluLeuArgHisThrProGluLeuPro  1440
Db  4261  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4320

Qy  1441  LeuAlaProArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysThrTyrAsp  1460
Db  4321  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4380

Qy  1461  ThrGlyAlaSerThrThrGlySerLysHisAspValArgSerLeuIleGlySerPro  1480
          ||||||||||||||||||||||||||||||||||||||||
Db  4381  NNNNNNNCGTCCACCACTGCCTCCAAAAAGCACACGACGTACGGTCCCTCATCGGCAGCCCC  4440

Qy  1481  GlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGlu  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4441  GGCCGGACGTTCCCACCCGTGCACCCGTCTGGATGATGGCCGACGCCCGGGCACTGGAA  4500

Qy  1501  ArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGly  1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4501  CGTGCCTGCTACGAGGAGAGCCTGAAGAGCCGCCCAGGGACCGCCAGCAGCTCGGGGGGC  4560

Qy  1521  SerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerPro  1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4561  TCCATTGCCCGCGGCGCCCCCGTCATTGTGCCTGAGCTGGGCAAGCCGCGGCAGAGCCCC  4620

Qy  1541  LeuThrTyrGluAspHisGlyValAlaProPheAlaGlyHisLeuProArgGlySerProVal  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4621  CTAACCTATGAGGACCACGGCGCACCCTTTGCCGGGCACCTCCCACGAGGTTCGCCCGTG  4680

Qy  1561  ThrMetArgGluProThrProArgLeuGlnLeuGluGlySerLeuSerSerLysAlaSer  1580
          ||||||||||||||||||||||||||||||||||||||
Db  4681  ACCACGCGGGAGCCCACGCCGCGGCCTGCAGGAGGNNNNNNNNNNNNNNNNNNNNNN  4740

Qy  1581  GlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrVal  1600
Db  4741  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4800

Qy  1601  ProGluHisProHisProIleSerProTyrGluHisSerLeuLeuArgGlyValSerGly  1620
Db  4801  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4860

Qy  1621  ValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGly  1640
Db  4861  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4920

Qy  1641  IleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThr  1660
Db  4921  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  4980

Qy  1661  TyrProHisLeuTyrProProTyrLeuLeuArgGlyTyrProAspThrAlaAlaLeuGlu  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  4981  TACCCGCACCTGTACCCCTACCTGCTGCGGGGCTACCCGGACACGGCGGCCCTGGAG  5040

Qy  1681  AsnArgGlnThrIleLeuAsnAspTyrIleThrSerGlnIleMetHisHisAsnThrAla  1700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5041  AACCGGCAGACCATCCTCAATGACTACATCACCTCGCAGATGATGCACCACAACACGGCC  5100

Qy  1701  ThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerSerLeu  1720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5101  ACCGCCATGGCCCAGCGGGCCGATATGCTGAGGGGCCTCTCGCCCCGAGTCTCGCTG  5160

Qy  1721  AlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleAspLeuSerGlnValProHisLeu  1740
```

```
Db  5161 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
         GCACTCAACTACGCTGCGGGTCCCGGGCATCATGCAGTGCCCAGTGTCCACACCTG  5220
Qy  1741 ProValLeuValProThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyr  1760
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5221 CCTGTGCTCGTGCCCCGACACCAGGCACCCAGCCACCGCCATGGACCGCCTGCCTAC  5280

Qy  1761 LeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSerProGly  1780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5281 CTCCCCACCGCGCCCCAGCCCTTCAGCAGCCGCCACAGCTCCCCACTCTCCCCAGGA  5340

Qy  1781 GlyProThrHisLeuThrLysProThrThrSerSerSerGluArgGluArgAspArg  1800
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5341 GGTCCAACACACTTGACAAAACCAACCACCAGCTCCTCGTCCGAGCGGGAGCGAGACCGG  5400

Qy  1801 AspArgGluArgAspArgAspArgGluArgGluLysSerIleLeuThrSerThrThr  1820
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5401 GATCGAGAGCGGGACACCGGGATCGGAGACGGGGAAAAGTCCATCCTCACGTCCACCACG  5460

Qy  1821 ValGluHisAlaProIleTrpArgProGlyThrGluLeuSerSerGlySerGlySer  1840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5461 GTGGAGCACGCACCCATCTGGAGACCTGGTACAGAGCAGCGGCAGCGGCAGC  5520

Qy  1841 SerGlyGlyGlyGlySerSerSerArgProAlaSerHisAlaHisHisGlnHis  1860
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5521 AGCGGCGGGGGTGGGGGCAGCAGCAGCCGCCCCGCCTCCCACGCGCACCACCAGCAC  5580

Qy  1861 SerProIleSerProArgThrGlnAsnAlaLeuGlnGlnAsnArgProSerValLeuHisAsn  1880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5581 TCGCCCATCTCCCCGACCCAGAATGCCCTCCAGCAGAACCCAGGTGCCTCACAAC  5640

Qy  1881 ThrGlyMetLysGlyLysIleIleThrAlaValGlyProSerLysProThrValLeuArgSer  1900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5641 ACAGGCATGAAGGGTATCATCACCGCTGTGGGCCCAGCACGCACGTCCTGAGGTCC  5700

Qy  1901 ThrSerThrSerProValArgProAlaAlaThrPheProProAlaThrPheHisCysProPro  1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5701 ACCTCCACCTCCCCGTTCGCCGCCGGTTGCCCACCATTCCCACCTGCCACCTGCCCA  5760

Qy  1921 LeuGlyGlyThrLeuAspGlyGlyValThrTyrProThrLeuMetGluProValLeuProLys  1940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5761 CTGGGGCGGCACCCTGATGGGGGTCTACCCTACCCTCATGGAGCCCGTTCTGCCCAAG  5820

Qy  1941 GluAlaProArgValAlaAlaArgProArgAlaArgProAlaArgAlaAspThrGlyHisAlaPheLeu  1960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5821 GAGGCCCCGCGGGTCGCCGCCCGGCCCCGAGCGCAGCACCGCCCATGCCTTCCTC  5880

Qy  1961 AlaLysProAlaAlaArgSerGlyLeuGluProAlaSerSerProSerLysProGlySerGlu  1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5881 GCCAAGCCCGCAGCCCGTCCGGGCTGGAGCCTGCCTCCCCAGCAAGGGCTCGGAG  5940

Qy  1981 ProArgProLeuValProProValSerProGlyHisGlyAlaThrArgThrProAlaLys  2000
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  5941 CCCCGGCCCCTAGTGCCTCCTGTCTCTGCCGGACACGGGGCCACCCGCACCCCGGCAAAG  6000

Qy  2001 AsnLeuAlaProHisHisAlaSerProAspProAlaSerProAlaSerAlaSerAspAsp  2020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6001 AACCTCGCACCTCACCACGCCAGCCCGGACCCCGCGCCCCGCCTCGGCCTCCGAC  6060

Qy  2021 ProHisArgGluLysLysThrGlnSerLysProPheSerIleLysProSerLysGluLeuArgSer  2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6061 CCGCACCGGGAAAAGAAGACTCAAAGTAAACCCTTTTCCATCAAGCCAAGTGGAACTGGAACTCCGTTCT  6120

Qy  2041 LeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerProValSer  2060
         |||    
Db  6121 CTGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6180

Qy  2061 SerProSerLeuThrHisAspLysLysGlyLeuHisLeuGluLeuGluAspLysSerSer  2080
         
Db  6181 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6240

Qy  2081 HisLeuGluGluGlyLeuLeuArgProLysGlnProGlyValLysLeuLysGlyGlyGlyGlyAla  2100


Db  6241 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCCCNNG  6300
Qy  2101 AlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerProLeuLeu  2120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6301 GCCCACCTCCCACACCTGCCGGCCCGCCTGCCTGAGACGCCAGCCCTCGTCCAGCCCGTGCTC  6360

Qy  2121 GlnThrAlaProGlyValHisGlyAsnArgValValThrLeuAlaGlnHisIleSer  2140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6361 CAGACCGCCCCAGGGGTCAAGGTCACCAGCGGGGTGGTCACCCTGCCCCAGCACATCAGT  6420

Qy  2141 GluValIleThrGlnAspTyrThrArgHisHisProGlnLeuSerAlaProLeuPro  2160
         
Db  6421 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6480

Qy  2161 AlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProPro  2180
         
Db  6481 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6540

Qy  2181 SerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySerProHisSer  2200
         
Db  6541 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6600

Qy  2201 GluGluGlyLysLysArgSerProGluProAsnLysLysProAsnLysSerValLeuGlyGlyGluAsp  2220
         |||||||||||                                          |||||||||||||
Db  6601 NNNNNNNNNNNNNNNNNNGTCTTGGGTGGTGAGGAC  6660

Qy  2221 GlyIleGluProValSerProProGluGlyMetThrGlyProGlyHisProGlyHisSerSerArgSerAla  2240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6661 GGTATTGAACCTGTGTCCCCACCCGAGGGCATGACGGGACCCAGGGCACTCCCGGAGTGCT  6720

Qy  2241 ValTyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys  2260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6721 GTGTACCCCGCTGCTACCGGGGATGGGAACAGAGACGGGCCCAGCAGGATGGGCTCCAAG  6780

Qy  2261 SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer  2280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6781 TCTCCAGGCAACACCAGCCAGCCTCCTGCACCTTCTTCAGCAAGCTGACCGAGAGCAACTCC  6840

Qy  2281 AlaMetValLysSerLysLysGlnGluIleLeuAsnLysLysGlnGluHisLysHisAsnArgAsn  2300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  6841 GCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGGAAT  6900

Qy  2301 GluProGluTyrAsnIleSerGlnInProGlyThrGluIlePheAsnMetProAlaIleThr  2320
         |||||||||||||||                                          
Db  6901 GAGCCTGAATACAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6960

Qy  2321 GlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMet  2340
         
Db  6961 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  7020

Qy  2341 GlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSer  2360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7021 GGAAGGCCAAGGTCTCTGGCAGACCCAGCAGCCGAAAGCCAAGTCCCCGGGGC  7080

Qy  2361 ProProLeuSerArgAlaAsnAlaPheAsnProLeuAlaSerAlaSerLeuProAlaAla  2380
         
Db  7081 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  7140

Qy  2381 MetProIleThrAlaAlaAspGlyArgSerAspHisThrLeuThrSerProGlyGly  2400
         
Db  7141 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGCGGC  7200

Qy  2401 GlyLysAlaLysValSerGlyArgProSerSerArgLysAlaLysSerProAlaProGly  2420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7201 GGAAGGCCAAGGTCTCTGGCAGACCCAGCAGCCGAAAAGCCAAGTCCCCGGGGC  7260

Qy  2421 LeuAlaSerGlyAspArgProProSerValHisSerSerGluGluValGlyAspCysAsn  2440
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7261 CTGGCATCTGGGGACCGGCCACCCTCTGTCCATTCCAGTGCAGTGGAGGAGGTGGGAGACTGCAAC  7320

Qy  2441 ArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAla  2457
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  7321 CGCCGGACACGCCGCTCACCAACCGGGTGTGGGAGGACAGGCCCTCGTCCGCA  7371
```

RESULT 2
AY412688

LOCUS   AY412688   7013 bp   DNA   linear   GSS 16-DEC-2003
DEFINITION   Mus musculus NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence,
             genomic survey sequence.
ACCESSION   AY412688
VERSION   AY412688.1 GI:39768653
KEYWORDS   GSS.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 7013)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,
             Adams,M.D. and Cargill,M.
  TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
             gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
  PUBMED   14671302
REFERENCE   2 (bases 1 to 7013)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sinsky,J.J.,
             Adams,M.D. and Cargill,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
             Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
             them based on alignment.
FEATURES   Location/Qualifiers
     source   1..7013
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene   <1..>7013
                     /gene="NCOR2"
                     /locus_tag="HCM4640"
ORIGIN

Alignment Scores:
Pred. No.:   3.86e-284
Score: 7109.50
Percent Similarity: 63.77%
Best Local Similarity: 61.37%
Query Match: 53.80%
DB: 29

Length: 7013
Matches: 1509
Conservative: 59
Mismatches: 768
Indels: 125
Gaps: 17

US-09-522-753-5 (1-2517) x AY412688 (1-7013)

Qy   1 MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrGluProArgTyrPro 20
Db   1 ATGTCAGGATCCACACAGCTCGTGGCACAGACAATGGCGGGCCTGCTGAGCCCGCTACCCA

Qy   21 ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu 40
                                    ||||||||||||||| ::||||||||:||||||
Db   61 CCCCATGGCCATCTCCTACCCGGTGCAGATAGCCCGGACGCACACGGACGTGGGGCTGCTT 120

Qy   41 GluTyrGlnHisHisSerArgAspTyrAlaSerHisArgLeuProGlySerIleIleGln 60
                      |||||||| |||||||||||||||||||
Db 121 GAGTACCAACCACCCCCCGTGACTACACCTGTCACCCGGTTCCATCATCCAG 180

Qy   61 ProGlnArgArgProSerLeuLeuSerGluSerGluPheGlnProGlyAsnGluArgSerGln 80
                                   ||||||||||||||||||||||:|||||||||
Db 181 CCACAGAGGAGGCCGCCCCTCACTGTGTCAGAGTTCCAGCCTGGGAGTGAACGNNNNNNNN 240

Qy   81 GluLeuHisArgProGluSerHisSerTyrLeuProGlyLeuProLeuGlyLysSerGluMet 100
Db 241 NNNNNNNNNNNNNNNNN-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 299

Qy 101 GluPheIleGluSerLysArgProArgLeuGluLeuProAspProLeuLeuArgPro 120

Db 300 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 359

Qy 121 SerProLeuLeuAlaThrGlnLeuGlyProAlaGlySerGluAspLeuThrLysAspArgSer 140
Db 360 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACCGTAGC 419

Qy 141 LeuThrGlyLysLeuGluProValSerProSerProProHisThrAspProGluLeu 160
                    ||||||||||||||||||||||||||||||||||||||||||||
Db 420 CTGGCAGGCAAGCTGGAGCCTGTGTCACCTTCACCTCCCCAGTCCCCGCACGCTGACCCTGAGCTA 479

Qy 161 GluLeuValProProArgLeuSerLysSerGluGlnLeuLeuGlnAsnMetAspArgValAsp 180
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db 480 GAGCTGGCGCCATCTGACGTTCCAGACGAGGAGGCCGCTGATCCAGAACATGGACCGCGTGGAC 539

Qy 181 ArgGluThrMetValGlnGlyGlnIleSerLeuLysLysLysGlnLeuGlnLeu 200
                ||||||||||||||||||||:|||||||||||||||||||||||||
Db 540 CGTGAGATCACCATGGTAGCAGCAGATCTCCAAGCTGAAGAAGAAGCAGCAACAGTTG 599

Qy 201 GluGluGlnAlaAlaLysProProGluLysProGluLysSerProProProIleGlu 220
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db 600 GAGGAGGAGGCCGCCAAGCCGCCGGAGAAACCCGAACCCGAGAAGTCTGTGCCGCCACCCATAGAA 659

Qy 221 SerLysHisArgSerLeuValGlnIleLleIleGluTyrArgAspGluAsnArgLysLysAlaGluAla 240
                |||||||||||||||||||:||||||||||||||||||||
Db 660 TCAAAGCACCGAAGCCTGGTCCAGATCATCTCCAGATCGAGAACCGGNNNNNNNNNNNN 719

Qy 241 AlaHisArgIleLeuGluGlyLyProGlnValGluLeuProLeuLeuTyrAsnGlnPro 260
Db 720 NNNNNNNNNNNNNNNNNNNNNNNCCCTCTGTACAACCAGCCG 779

Qy 261 SerAspThrArgGlnTyrHisGluAsnGlnIleLysIleLeuAsnGlnAlaMetArgLysLeuLeu 280
                    |||||||||||||||||||||||||||||||||||||||||||||||
Db 780 TCTGACACACGCCAGTACCATGAAAATCAAAATAAACCAGGCATGCGGAAGAAGCTG 839

Qy 281 IleLeuTyrPheLysArgArgAsnHisAlaAlaArgLysSerGlnTrpLysSerPheCysGln 300
                                   :::|||:::::|||||:|:
Db 840 ATCTTGTACTTTAAGCGGAGGAACCACGCGGCCAGCCGCCGTGGAACAGCCGTCTCCGCCAG 899

Qy 301 ArgTyrAspGlnLeuMetGlnLeuGluGlyLeuLysLysSerValGluArgIleGluGluAsnAsnPro 320
Db 900 CGTATGACCAGCTCATGGAGCGGTGGGAGAAAGGTAGAGCGCGCATAGAGACAATCCG 959

Qy 321 ArgArgArgAlaAlaLysGluGluLysSerValSerGluLysTyrTyrGluLeuProGluIle 340
                     ||||||||||||||||||||:|||||||||||||||||||||:|||||
Db 960 CGAAGGAGGGGCCAAGGAGGAGAAGAGTGTGAGTGAGTACTACGAGAAAACAGTTCCCGGAGATC 1019

Qy 341 ArgLysGlnArgGlyLeuGlnLeuGlyArgMetGlnSerArgValGlyArgGlySerGly 360
                ||||||||||||||||||||:|||||||||||||||||||||||||||
Db 1020 CGCAAGCAGCGCGGAGCGTGCAGAGCAGCCGCATGCAGAGCAGGGTGGGCCGTGGCAGTGGG 1079

Qy 361 LeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
                ||||||||||||||||||||:||||||||||||||||||||||||||
Db 1080 CTCTCCATGTCGGCTGCCGCCGTCAGTGAGCATGAGGTTTCTGAGATCATTGATGGCTTGTCT 1139

Qy 381 GluGlnAsnLeuGlyLysGluGlnArgGlnLeuGlnMetAsnMetTrpSerTyr 400
                |||||||||||||||:||||||||::|||||||:|||||||
Db 1140 GAGCAGAACCTGGGCAAGGAGCAGAGGCAGCTGCAGATGAACATGTGGAGCTACCTGTAC 1199

Qy 401 AspAlaAspGlnAsnArgIleLysPheAsnGlyLeuMetAlaAsnProMet 420
                |||||||||||||||||||:||||||||||||||||||||
Db 1200 GACGCGGACCAGAACCGCATCAAGTTCATCAACATGATGGCCAATCCCATG 1259

Qy 421 LysValTyrLysAspArgArgGlnLnValMetAsnMetTrpSerGluGlnLysGluThrPhe 440
                                   :::::||||||||||:
Db 1260 AAGGTCTACAAGGACCGTCAGGTTACCAACATGTGGAGCGAGCAGAAGGGAGACACCTTC 1319

Qy 441 ArgGluLysPheMetGlnHisProLysProLysAsnPheGlyLeuIleAlaSerPheLeuArg 460
                ||||||||||||||||||||||||||||||||||||||||
Db 1320 CGTGAGAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1379

Qy 461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLys 480

```
Qy  841 GluGluGlnLysProProAlaAlaGluGluLeuAlaAlaValAspThrGlyLysAlaGluGlu 860
            ::  .::::::::|||...
Db 2286 AGCCCAGGAGGACCCGAGGCCTGAGGCG                                     2324

Qy  861 ProValLysArgGluCysThrGluGluAlaGluGlyProAla---LysGlyLysAsp        879
            |||  ||||||||||  ||||  |||
Db 2325 AGTGTAAAGAGTGACCACAAGGAGGAGACCGAGGAAGAGCAAAGCCAAGGGC            2384

Qy  880 AlaGluAlaAlaGluThrAlaGluGlyAlaLeuLysAlaGluLysLysGluGlyGly        899
           ||||  :::|||
Db 2385 ACAGAGGGCCATTGAAACTGTCTGAGGCACCACTTAAGGTGGAG------GAGGCTGGT      2438

Qy  900 SerGlyArgAlaThrThr---AlaLysSerGlyAlaProGlnAspSerAspSer          918
            ||||||
Db 2439 AGCAAGGCAGCTGTGACCAAGGGTTCCAGCTCAGGTGCCACCCAGGACAGTGACTCCAGT    2498

Qy  919 AlaThrCysSerAlaAspGluValAspGluGluGlyGlyGlyAspLysAsnArgLysLeuLeu 938
            |||||
Db 2499 GCCACCTGCAGTGCCGATGAGGTGGACGAACCACAAGGGCAGGCTGCTG               2558

Qy  939 SerProArgProSerLeuThrProThrGlyAspProArgAlaAsnAlaSerProGln       958
            ||||||
Db 2559 TCACCAAGGCCCAGCTCCTCACCCGGCTGGAGATCCCGGGCCAGTACCTCGCCCCAG       2618

Qy  959 LysProLeuAspLeuLysGlnLeuLysGlnArgGlnLeuAlaAlaAlaIleProProIleGlnIleVal 978
            ||||||
Db 2619 AAGCCGCTGAAGCTGAAGCAGCTGAAGCAGCGGCAGCTGGCCGCCGCCATCCCCCCTATCCAGGTC 2678

Qy  979 ThrLysValHisGluProProArgGluAspAlaAlaProThrLysGlnProValProProAla 998
            ||||||
Db 2679 ACCAAGGTCCATGAGCCCCCGCGGGAGGACGCAGCCCCCAAGCCAGTTCCCCCTGTG       2738

Qy  999 ProProProProAsnLeuGlnProGluSerAspAlaProGlnGlnProGlySerSer      1018
            ||||||
Db 2739 CCTCCACCCACGCAACCTACAGCCCCAGAGGGTGCTCAGCAGCCTGGGAGGAAGT         2798

Qy 1019 ProArgGlyLysSerArgSerProAlaProAlaAspLysProLysAlaPheAlaAlaGlu   1038
            ||||||  :::|||
Db 2799 CCACGTGGCAAGTCCCGCAGTCCCGCCCCAGCCGAGAAACCCAAGGCGNNNNNNNNNN      2858

Qy 1039 AlaGlnLysLeuProGlyAspProCysTrpThrSerGlyLeuProProPheProValPro   1058
Db 2859 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN             2918

Qy 1059 ProArgGluValIleLeuLysAlaSerProHisAlaProAspProSerAlaPheSerAla   1078
Db 2919 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN             2978

Qy 1079 ProProGlyHisProLeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArg   1098
Db 2979 NNNNNNNGTCACCCGCTGGCCTCCACGATAGTGCCCGTCCTGCCACGT                3038

Qy 1099 ProProThrIleSerAsnProProLeuIleSerSerAlaLysHisProSerValLeu     1118
Db 3039 CCCCCC---ATCTCTAACCCCCCACCCCTCATCTCCTCTGCCAAGCATCCCGGCGTACTT    3095

Qy 1119 GluArgGlnIleGluGlyAlaIleSerGlnLeuMetSerValGlnLeuHisValProTyrSer 1138
            |||  :::|||
Db 3096 GAGAGGCAGATCGAGGGTGCCATCTCCCAGGGGATGTCAGTCCAGCTTCGTGTGCCCACTCA 3155

Qy 1139 GluHisAlaLysAlaProValThrMetGlyLeuProLeuProMetGlyAspPro        1158
Db 3156 GAGCATGCCAAGGCCCCAGTGACCATGGGGCTGCCCCTTGCCATGGGACCCT            3215

Qy 1159 LysLysLeuAlaProPheSerGlyValLysGlnGlnLeuSerProArgGlyGlnAla     1178
            |||  :::|||
Db 3216 AAGAAGCTGGCCCCCTTCAGTGAGTGAAGCAGGAACAGTTGTCCCCTGGGGTCAGGCT     3275

Qy 1179 GlyProProGluSerLeuGlyValProThrArgAlaGlnThr                    1198
            :::|||
Db 3276 GGGCCACCTGAAGTCTGGGAGTGCCCACTGCTCAGGAGGACCCTGTGCTGAGAGGGGACA   3335
```

```
Db 1380 NNNACGGTCCGCTGAGTGTGTCCTCTATTACTACCCTGACCAAGAAGAATGAAAATTACAAG 1439

Qy  481 SerLeuValArgSerTyrArgArgGlyLysSerGlnGlnIleGlnGlnGlnGln          500
            ||||||||||||
Db 1440 AGCTTGGTTGAGGCGGAGCTATCCGCGCCGTGGCAAGAGCCAGNNNNNNNNNN           1499

Qy  501 GlnGlnGlnGlnGlnGlnGlnProMetProArgSerGlnGluGluGluLys            520
Db 1500 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN                 1559

Qy  521 AspGluLysGluLysGluGluAlaGluGluLysGluGluGluGluGluLysGluAsn      540
Db 1560 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN                 1619

Qy  541 AspLysGluAspLeuLeuLysGluLysThrAspAspThrSerGlyThrGluAspAspGlu   560
            |||||
Db 1620 NNNGGAGAAGAACAGACAGACACACTTCTGCGAGGACAACGATGAG                 1679

Qy  561 LysGluLeuAlaValAlaSerLysLysGlyArgLysThrAlaAsnSerGlnArgArgLysGlyGly 580
            |||||
Db 1680 AAAGAGCTGGCCGTGGCCTCCAAAGGCCGCAAAACTGCCAACAGCCCAAGGCCGCCCAAAGGC 1739

Qy  581 ArgIleThrArgSerMetAlaAsnGluLeuAlaAsnSerGluGluValGluLeuAlaIleThrProGlnGln 600
            |||||
Db 1740 CGTATCACGCGCTCCATGGCCAACGAGGCCAACCATGAGGAGCAGCCACCACAGCAA      1799

Qy  601 SerAlaGluLeuAlaSerMetGluLeuAsnGluSerSerArgTrpThrGluGluMet      620
            :::|||
Db 1800 AGTTCAGGATGCTGGCTTCCATGGAGTAACGAGAGTTCTCGCTGGACTGAGGAAGAGATG   1859

Qy  621 GluThrArgAlaLysLysGlyLeuLeuGluHisGlyArgAsnTrpSerAlaIleAlaArgMet 640
            |||||
Db 1860 GAGACACAGGCAAAGAAAGGCCTCCTGGAACATGGGAACTGGTCAGCCATTGCCCGCATG   1919

Qy  641 ValGlySerLysThrValSerGlnCysLysLysAsnPheTyrPheAsnPheTyrLysLysArgGln 660
            |||||
Db 1920 GTGGGCTCCAAGACCGTGTCCCAGTGTAAGAAGAACTTCTACTTCAACTACAAGAAGAGACAG 1979

Qy  661 AsnLeuAspGluIleLeuGlnGlnHisLysLeuLeuMetGluLysGluGluArgAsnAlaArg 680
            |||||
Db 1980 AACCTGGACGAAATCCTTCAGCAGCACAAGCTAAAGATGGAGAAGGGAGGAACGCTCGG    2039

Qy  681 ArgLysLysLysLysAlaProAlaAlaAlaSerGluGluAlaAlaSerGluLysAlaLeuAlaProProValVal 700
            |||||
Db 2040 AGGAAGAAGAAGAAGGCCCAGCTGCGGCGAGCGAGGACGACCCCTTCCCACCTGCCGCT   2099

Qy  701 GluAspGluLeuMetGluGluSerGlyValSerGlyValSerGlyAsnGluValSerGluGlu 720
            |||||
Db 2100 GAGGACGAAGAGATGGAAGCATCAGGCGCCAGTGCCAATGAGGAAGAGCTGGCGGAGGAG   2159

Qy  721 AlaGluAlaLeuHisAlaSerGlyAsnGluValProArgGlyValGluCysSerGlyProAla 740
            |||||
Db 2160 GCAGAAGNNNNNNNNNNNNNNNN---------------------------------NNN   2189

Qy  741 ThrValAsnAsnSerSerAspThrGluSerIleProSerProHisThrGluAlaAlaLys   760
            :::|||
Db 2190 NNNNNNNNNNNNNNNNNNNNNNNNNCTGCCCAGCTCCCCAGACACAGAGATGCCAAG      2249

Qy  761 AspThrGlyGlnAsnGlyProLysProAlaThrLeuGlyAlaAspGlyCysSerGlyProPro 780
            :::
Db 2250 GCAGAAGNNNNNNNNNN--------------------------------------------  2252

Qy  781 GlyProProThrProArgArgThrSerArgArgAlaProIleGluGluProThrProAlaSer 800
Db 2252 ------------------------------------------------------------   2252

Qy  801 GluAlaThrGlyAlaProThrProProProSerProSerAlaProProPro           820
Db 2252 ------------------------------------------------------------   2252

Qy  821 ValValProLysGluGluLysGluGluLeuThrAlaAlaAlaAlaProProValGluGluGly 840
            :::||
Db 2253 ------CAGAAGTCTGAGGCCGAGGAG----------------------ATCGATGTGGGA   2285
```

```
Qy  1199 AlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThrArgValPro 1218
Db  3336 GCACTGGGCTCCGCACCAGTGAAGCATCACCAAGGGCCTCCCCAGTACCCGGGGCTGCA 3395

Qy  1219 SerAspSerAlaIleThrTyrArgGlySerIleThrProAlaAspValLeu 1238
Db  3396 GACGGCCCC------AGCTACAGAGAGCTCTATCACCCCAGCCACGCCCGGAGACGTCCTC 3449

Qy  1239 TyrLysGlyThrIleThrArgIleIleGlyGluAspSerProSerArgLeuAspArgGly 1258
Db  3450 TACAAGGGTACCATCACCAGCAGGATCGTCGGTGGAGACAGCCCAAGTCGCCTTGACCGGCA 3509

Qy  1259 ArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLyLysGlyHisValLeu 1278
Db  3510 CGAGAGGAGACACCCCTGCCCCAAGGGCCATGTCATCTATGAGGGCAAGAAAGGCCACGTCCTA 3569

Qy  1279 SerTyrGlyGlyMetSerValThrGlnCysSerLysGluAspGlyArgGlySerSerSer 1298
Db  3570 TCCTATGAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3629

Qy  1299 GlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetGluGlyArgVal 1318
Db  3630 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3689

Qy  1319 GlyArgAlaIleSerSerAlaSerIleGluGluGlyLeuMetGlyArgAlaIleProProGlu 1338
Db  3690 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGACTCATGGGCCGCGGCCATC---CCTGAG 3746

Qy  1339 ArgHisSerProHisHisLeuLysGluGlnHisHisIleGlyGlySerIleThrGlnGly 1358
Db  3747 CAGCACCAGCCCC---CACCTCAAGGAGCAGCATCACATCGGAGGCTCCATCACGCAAGNN 3803

Qy  1359 IleProArgSerTyrValGlnLaGlnGluTyrLeuAspTyrLeuAlaLysLeuLeu 1378
Db  3804 NNNNN-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3862

Qy  1379 LysArgGluGlyThrProProProProProSerArgAspLeuThrGluThrGluTyrLys 1398
Db  3863 NNNNNN------------------NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3895

Qy  1399 ThrGlnAlaLeuGlyProLeuLeuLysProAlaHisGluGlyLeuValAlaThrVal 1418
Db  3896 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3955

Qy  1419 LysGluAlaGlyArgSerIleHisGluLeuProArgGluLeuProArgGluHisThrProGlu 1438
Db  3956 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4015

Qy  1439 LeuProLeuAlaProArgProLeuLysGlyLySerIleThrGlnThrProLeuLys 1458
Db  4016 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4075

Qy  1459 TyrAspThrGlyAlaSerThrThrGlySerLysLysLysHisAspValArgSerLeuIleGly 1478
Db  4076 NNNNNNNNNNNNNCACCCTCCACTGGCACAGAAACACGACGGTGCCTCCATCATGGC 4135

Qy  1479 SerProGlyArgThrPhePheProProValHisProLeuAspValMetAlaAspAlaArgAla 1498
Db  4136 AGCCCCGGCCGGACCTTTCCCTGCCCTGACACCATAATGGCTGATGCCCCGGCA 4195

Qy  1499 LeuGluArgAlaCysTyrGlyGluThrSerArgProGlyProThrAlaSerSerSer 1518
Db  4196 CTGGAGCGTGCCTGCTATGGAGAGACTCAGAGGACCCGTCAGGGACCAGTGGTGCA 4255

Qy  1519 GlyGlySerIleAlaAlaProValIleValAlaProValHisProGluLeuGlyLysProArgGln 1538
Db  4256 GGGGGCTCCATCACACGTGGGCTCCAGTCGTCGTGGCGTGGCAAGCCACGCGCCAA 4315

Qy  1539 SerProLeuThrTyrGluAspHisHisIleAlaProPheAlaGlyHisIleProArgGlySer 1558
Db  4316 AGCCCACTGACCTTACGAAGACCACGGGCACCCTTCACCACGTCCACGTGGCTCC 4375

Qy  1559 ProValThrMetArgGluProThrProArgLeuAspArgLeuGlnGluGlySerLeuSerLys 1578
```

```
Db  4376 CCTGTGACCACGAGGGAGGCCCACGCACCACGCCTTCAGGAAGNNNNNNNN 4435

Qy  1579 AlaSerGlnAspArgLysThrSerThrProArgGluIleAlaLysSerProHisSer 1598

Db  4436 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4495

Qy  1599 ThrValProGluHisHisProIleSerProTyrGluHisLeuLeuArgGlyVal 1618

Db  4496 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4555

Qy  1619 SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638

Db  4556 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4615

Qy  1639 ArgGlyIleProLeuAspAlaAlaAlaAlaAlaTyrLeuProArgHisLeuAlaProAsn 1658

Db  4616 NNNNNNNNNNNNNNNNNNNCCGAGCCTACTTACCTGCCCCGGCACTGGCCCCAGC 4675

Qy  1659 ProThrTyrProHisLeuTyrProLeuLeuArgGlyTyrProAspThrAlaAla 1678

Db  4676 CCCACCTACCCACACCTGTACCCACCTTACCTCATCCCGGCTACCCTGACACGGCGGCC 4735

Qy  1679 LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnMetHisHisAsn 1698

Db  4736 CTGGAGAACCGCCAGACACTCATCAATCATGACTACATCACCTCGCAGATGCACCACAAC 4795

Qy  1699 ThrAlaThrAlaMetAlaGlnAspMetLeuArgGlyLeuSerProArgGluSer 1718

Db  4796 GCTGCCTCCGGCCAGCGTGCTGACATGCTGAGGGGTCGTCACCGGAGAGTCC 4855

Qy  1719 SerLeuAlaLeuAsnTyrAlaAlaGlyLeuProArgGlyIleIleAspLeuSerGlnValPro 1738

Db  4856 TCGCTGGCCCTCAATTATGCGGTGGCCCAAGAGACGATTATCGACCTGTCCAAGTGCCA 4915

Qy  1739 HisLeuProValLeuValProThrProProGlyThrProProAlaThrAlaMetAspArgLeu 1758

Db  4916 CACCTGCCCGTCGGTGGTGCCCAAGCGCACCCGGCACCCCGCCACGGCAATGACCACCCCTT 4975

Qy  1759 AlaTyrLeuProThrArgAlaProGlnProPheSerSerSerArgHisSerSerProLeuSer 1778

Db  4976 GCCTACCTTCCCCACTGCGCCCCCCCTTCAGCAGCGCGCCACAGTAGCTCACCGTGTCC 5035

Qy  1779 ProGlyGlyProThrHisLeuThrLysProSerSerSerGluArgGluArg 1798

Db  5036 CCAGGAGGCCCCACTCACCTAGCTAAACCATCTTCATCGGAGCGGGAACGG 5095

Qy  1799 AspArgAspArgGluArgAspArgAspArgGluGluArg 1818

Db  5096 GAACGTGAGCGGGAACGAGAC------------AAGTCCATCCTCACGTCTACC 5137

Qy  1819 ThrThrValGluHisAlaProIleTrpArgProGlyThrProGlyGlnSerGlySerSer 1838

Db  5138 ACTACAGTCGAGCATGACCCCATCTGGAGACCTGGTACGGACAGAGCGGGGCT--- 5194

Qy  1839 GlySerSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlySerSerArgProAlaSerHisAlaHis 1858

Db  5195 ----------------------GGGGGGCAGCAGCCGCCCCTCCCACC------CAC 5227

Qy  1859 GlnHisSerProIleSerGlyIleIleThrAlaValGluProSerValLeu 1878

Db  5228 CAGCACTCGCCCCATCTCCCCCGGACACCGCCCAGGACGCCAGTGGTGCA 5287

Qy  1879 HisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrValLeu 1898

Db  5288 CACAACACGAGCATGAAGGGGTGTCACCTCGGTGACCGGCAACCCGGCACGGTCCTG 5347

Qy  1899 ArgSerThrSerThrSerProValArgProAlaArgProAlaAlaThrPheProProAlaThrHis 1918

Db  5348 AGGTCCACTTCCACCTCTTCGCCTGTCGCCCCAGTGCCACATTCCACCTGCCACCCAC 5407

Qy  1919 CysProLeuGlyThrLeuAspGlyValTyrProThrLeuMetGluProGluProValLeuLeu 1938
```

```
Db  5408 TGCCCACTTGGTGCACCCTGAAGGGTCTACCCTACCCTCATGGAGCCCGTCCTGTTA 5467
Qy  1939 ProLysGluAlaProArgValAlaArgProGluArgProArgAlaAspThrGlyHisAla 1958
              ||||| |  |||||     ||     ||
Db  5468 CCCAAGGAAGACCTTCTCGGGTCGCCCGGCCCGGTGGACGCTGGACCATGCC 5527
Qy  1959 PheLeuAlaLysProProAlaArgSerGlyLeuGluSerProSerLysGly 1978
              |||||||||||||  |
Db  5528 TTCCTCACCAAACCCCGCCCCG------GAGCCGCCTCCTCCACCAGCAAGAGC 5578
Qy  1979 SerGluProArgProLeuValProValSerGlyHisAlaThrIleAlaArgThrPro 1998
         ||||||  || |||      |||  ||
Db  5579 TCCGAGCCCCGATCCTAGCACCCCCAGCTCCAGCCACACAGCCATCGCCCGCACCCCA 5638
Qy  1999 AlaLysAsnLeuAlaProHisHisAlaSerProAspProProAlaProAlaSerAla 2018
              |||:||||||||:|||  |||         ||| ||||
Db  5639 GCAAAGAGCCTTGCACCCCACCCAGTCCGACCCCGCCGG---CCCACCTCGGCC 5695
Qy  2019 SerAspProHisArgGluLysThrGlnLysSerProPheSerIleGlnLeuGluLeu 2038
           |||  ||  ||||  |||| | |||     |
Db  5696 TCAGATCTGCACCGAGAAAAGACTCAAAGTAAACCCTTTCCATCCAGGAATTGGAACTC 5755
Qy  2039 ArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluProValSerPro 2058
              |||
Db  5756 CGTTCTCTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5815
Qy  2059 ValSerSerProSerLeuThrHisAspLysProLysHisGlyLeuGluLeuGluLeuAsp 2078
Db  5816 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5875
Qy  2079 LysSerHisLeuGluGlyLeuGluLeuArgProLysGlnProGlyLeuProLysLeuGly 2098
                                      ||:::||||||
Db  5876 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCCCCATGAAGCTCAGCGGCG 5935
Qy  2099 GluAlaAlaHisLeuProHisLeuProLeuArgProLeuProHisLeuProSerGlnProSerSerPro 2118
         |||||||||||||||||||
Db  5936 GAGGCTGCCCATCTCCCACATCTGCGGCCCACTGCCCAGAGCCCCTCATCCAGCCCA 5995
Qy  2119 LeuGluGlnThrArgAlaProGlyValLysLysGlyValHisArgGlnValThrLeuAlaValThrLeuAlaGlnHis 2138
            |||      ||:::|||||  ||||||
Db  5996 CTCCTCCCAGACTGCCCCAGGCATCAAAGTACCAGAGAGGTGTGCACCCTGGCTCAGCAC 6055
Qy  2139 IleSerGluValIleIleThrGlnAspTyrThrArgHisHisAspHisProGlnGlnLeuSerAlaPro 2158
          |||||
Db  6056 ATCAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6115
Qy  2159 LeuProAlaProLeuTyrSerPheSerPheProGlyProAlaSerCysProValLeuAspLeuArgArg 2178
Db  6116 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6175
Qy  2179 ProProSerAspLeuTyrLeuProProAspHisGlyAlaProAlaArgGlySerPro 2198
Db  6176 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6235
Qy  2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValGlyGlyGly 2218
                              ||:::||||||
Db  6236 NNNNNNNNNNNNNNNNNNGTCCCAGAACCAGCCAAAAACATCGGTCTGGGCAGCAGT 6295
Qy  2219 GluAspGlyIleGluProGluValSerProProGlnGlyMetThrGluProGlyGlyMetSerSerArg 2238
             |||||||          |||:||||
Db  6296 GAGGATGCCATTGAGCCTGTGTCCCCACCAGAGGGCATGACTGAGCCAGGACAGTGCTCGG 6355
Qy  2239 SerAlaValTyrProLeuGluTyrArgProLysTyrArgProGluValGluProSerArgMetGly 2258
         ||||      |||| |
Db  6356 AGCGCTGTGTACCCATCTGTTATCGAGACGGGGACAACAGGGCCAGCAGGATGGGC 6415
Qy  2259 SerLysSerProGlyAsnThrSerGlnProProAlaPheSerLysLeuThrGluSer 2278
              |||        ||||    |||||||| |
Db  6416 TCTAAGTCTCCAGGCAACACCAGCCAGCCTTCTTCAGTAAGCTGACTGACAGAGAGC 6475
Qy  2279 AsnSerAlaMetValLysSerLysLysGlnGluIleIleAsnLysLeuAsnThrHisAsn 2298
           ||||          |||| |||
Db  6476 AACTCCCGCATGTGTGAAGCTGAAGAAGAAGATCAACAAGAAAACTCAACACCCCAAC 6535
```

```
Qy  2299 ArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAla 2318
         ||  |||   ||
Db  6536 CGGAACGAGCCAGAATACANNNNNNNNN 6595
Qy  2319 IleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnGluHisHisAlaSerThr 2338
Db  6596 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6655
Qy  2339 AsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlu 2358
Db  6656 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6715
Qy  2359 GluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuPro 2378
Db  6716 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6775
Qy  2379 AlaAlaMetProIleThrArgAlaAlaAspGlyArgSerAspHisSerThrLeuThrSerProGly 2398
Db  6776 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT 6835
Qy  2399 GlyGlyGlyAlaAlaValSerGlyArgProSerSerArgLysAlaLysProSerProAla 2418
         |||||||
Db  6836 GGAGGTGGGAAAGCCAAGGTCTGGCAGCAGCCTAGCAGCCAAAGCCAAGTCGCCAGCA 6895
Qy  2419 ProGlyLeuAlaSerGlyLysAspArgProSerValSerSerValHisSerGluGlyLysAsp 2438
         |||||||||||||||
Db  6896 CCAGGCCTAGCGTCCGGAGACCGACCACCCCCTTCTGTCTCCTCAGTACACTCAGAGGGGGAC 6955
Qy  2439 CysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAla 2457
          ||  |||    |||  |||||
Db  6956 TGCAATCGCCGAACACCACTCACCAACCGTGTGTGGGAGGACCCGCCCTCATCTGCA 7012
```

RESULT 3
```
LOCUS       AY412687                6836 bp    DNA     linear   GSS 16-DEC-2003
DEFINITION  Pan troglodytes NCOR2 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY412687
VERSION     AY412687.1  GI:39768652
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 6836)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
  PUBMED    14671302
REFERENCE   2  (bases 1 to 6836)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..6836
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene            <1..>6836
                     /gene="NCOR2"
                     /locus_tag="HCM4640"
ORIGIN
```

```
Alignment Scores:
Pred. No.:          2.78e-260      Length:         6836
Score:              6533.00        Matches:        1407
Percent Similarity: 61.81%         Conservative:   1
Best Local Similarity: 61.76%      Mismatches:     870
Query Match:        49.44%         Indels:         0
DB:                                Gaps:           0
```

US-09-522-753-5 (1-2517) x AY412687 (1-6836)

```
Qy    27  ProValGlnIleAlaArgThrHisAspValGlyLeuLeuGluTyrGlnHisHisSer  46
          ||||||||||||||||||||||||||||||||||||||||||||
Db     2  CCAGTGCAGATTGCCCGGACGACGTCGGCGTCCTCGAGTACCAGCACCACTCC  61

Qy    47  ArgAspTyrAlaSerHisLeuSerProGlySerIleIleGlnProGlnArgArgPro  66
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  CGCGACCATGCCTCCCACCTGTCGCCCGGCTCCATCATCCAGCCGCAGCGGAGGCCC  121

Qy    67  SerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGlnGlnLeuHisLeuArgPro  86
          ||||||||||||||||
Db   122  TCCCTGCTGTCTGAGTTCCAGCCCGGAATGAACGNNNNNNNNNNNNNNNNNNN  181

Qy    87  GluSerHisSerTyrLeuProGlyLeuLysSerGluMetGluProIleIleGluSerLys  106
          |||||||||||||||||
Db   182  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  241

Qy   107  ArgProArgLeuGluLeuProAspProLeuLeuArgProSerProLeuLeuAlaThr  126
          ||||||||||||||||
Db   242  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  301

Qy   127  GlyGlnProAlaGlySerGluAspLeuThrLysAspArgSerLeuThrGlyTyrLysLeuGlu  146
          |||||||||||||||
Db   302  NNNNNNNNNNNNNNNNNNNNNNNAGCCTGACGGGCAAGCTGGAA  361

Qy   147  ProValSerProProHisThrAspProGluLeuGluLeuValProProArg  166
          |||||||||||
Db   362  CCGGTGTCTCCCCCAGCCCCCGCACACCGACCCTGAGCTGGTGCCGCCGCGG  421

Qy   167  LeuSerLysGluGluLeuGlnAsnMetAspArgValAspAspArgGluIleThrMetVal  186
          |||||||||||||||||||||||||||||||||||||
Db   422  CTGTCCAAGGAGGAGCTGCAGAACATGGACCGCGTGGACGACCGCGAGATCACCATGGTA  481

Qy   187  GluGlnGlnIleSerLysGluLeuLysLysGlnIleGlnLeuGluGlnLysAlaAlaLys  206
          |||||||||||||||||||||||||||||||||||||||
Db   482  GAGCAGCAGATCTCTAAGCTGAAGAAGCAGNNNNNNNNTGGAGGAGGAGGCTNCCAAG  541

Qy   207  ProProGluProGluProGluValSerProProIleGluSerProIleGluSerLysLeu  226
          ||||||||||||||||||
Db   542  CCGCCCGAGCCTGAGAAGCCCGTGTCACCGGCCATCGAGTCGAAGCAGCACCGCAGCCTG  601

Qy   227  ValGlnIleIleIleTyrAspLysLysArgLysAsnArgLysLysAlaAlaAlaHisArgIleLeuGlu  246
          ||||||||||
Db   602  GTGCAGATCATCTACGACGAAGAACCGGNNNNNNNNNNNNNNNNNNNNNNNNN  661

Qy   247  GlyLeuGlyProGlnLeuProLeuTyrAsnGlnProGlnThrArgGlnTyr  266
          ||||||||||||||||||||||||||||||||||||||||
Db   662  NNNNNNNNNNNNNNNNNCCGCTGTACAACCAGCCTCCACACCGACAGTAT  721

Qy   267  HisGluAsnIleIleLysIleLysAsnGlnAlaMetArgLysLysLeuTyrPheLysLysArg  286
          |||||||||||||||||||||||||||||||||||||||
Db   722  CATGAGAACATCAAAATAAACCAGGGCGATGCGGAAGAAGCTNAATCTTGTACTTCAAGAGG  781

Qy   287  ArgAsnHisAlaArgGlyGlnTrpLysTrpLysPheCysGlnArgTyrAspProGlnLeuMet  306
          |||::::|||||||||||||||||||||||||||||||||||||
Db   782  AGGAATCACGCTCGGAACAATGGGAGCAGAAGTTCTGCCAGCGTTATGACCCAGCTCATG  841

Qy   307  GluAlaLeuGluLysValArgIleGluValGluAsnAsnProArgArgArgAlaLysGlu  326
          |||||||||||||||||||||||||||||||||||||||
Db   842  GAGGCCCTGGAGAAGGAAGAAGTGAAGATGGAAGAAGAAGAAGAAGAAGAAAGCG  901

Qy   327  SerLysValArgGluTyrTyrGluLysGlnPheProGluIleArgLysGlnArgGluLeu  346
          |||||||||||||||||||||||||||||||
Db   902  AGCAAGGTGCGCGAGTACTACGAGAAGCAGTTCCCTGAGATCCGCAAGCAGCGCGAGCTG  961
```

(right column, continued)

```
Qy   347  GlnLeuArgMetGlnSerArgValGlyLeuGlySerGlyLeuSerMetSerAlaAla  366
          ||||||||||||||||||||||||||||||||||||||||||
Db   962  CAGAGACGCATGCAGAGACAGGGTGGCCCAGCGGGCGGGGGTGGGTGTCCATGTCCGCTGCC  1021

Qy   367  ArgSerGluHisGluValSerGlyValIleIleAspGlyLeuSerGluGlnLeuAsnLeuGlu  386
          ||||||||||||||||||||||||||||||||||||||||
Db  1022  CGCAGCGAGCACGAGGTGTCAGGAGTCATCATCGATGGCCTCTCAGAGCAGGAGAACCTGGAG  1081

Qy   387  LysGlnMetArgGlnLeuAlaValIleProLeuProProMetLeuTyrAspAlaAspProGlnGlnArg  406
          |||||||||||||||||||||||||||||||||||||
Db  1082  AAGCAGATGCGCCAGCTGGCCGTGATCCCGCTGCCTCGTATGCTGTACGACGCCGACCCAGCAGCGC  1141

Qy   407  IleLysPheIleLeuAsnMetAlaAsnGlyLeuMetAlaAspProMetLysValTyrrLysAspArg  426
          |||||||||||||||||||||||||||||||||||||
Db  1142  ATCAAGTTCATCAACATGACAGCGGTTATGGCTGACCCCATGAAGGTGTACAAAGACCGC  1201

Qy   427  GlnValMetAsnMetTrpSerGluGlnGluLysGluThrPheArgGluLysPheMetGln  446
          ||||||||||||||||||||||||||||||||||
Db  1202  CAGGTCATGAACATGTGGAGTGAGCAGGAGAAGGAGACCTTCCGGGAGAANNNNNNNNNN  1261

Qy   447  HisProLysPheGlyLeuIleAlaSerPheLeuGluArgLysThrValAlaAlaGluCys  466
          |||||||||||||||||||||||||
Db  1262  NNNNNNNNNNNNNNNNNNNNNNNGTGGCTGAGTGC  1321

Qy   467  ValLeuTyrTyrTyrLeuThrLysTyrLysAsnGluAsnTyrLysSerLeuValArgArgSer  486
          |||||||||||||||||||||||||||||
Db  1322  GTCCTCTATTACTACCTGACTAAGAACATGAGAACTATAAGACGCTGGTGAGACCGGAGC  1381

Qy   487  TyrArgArgArgGlyIleGlnGlnGlnGlnGlnGlnGlnGlnGln  506
          ||||||||||
Db  1382  TATCGGCGCCGCGGCAAGAGCCAGNNNNNNNNNNNNNNNNNNNNNNN  1441

Qy   507  GlnGlnGlnGlnMetProMetProArgSerGlnGluGluLysLysAspGluLysLysGlu  526
          ||||||||||
Db  1442  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  1501

Qy   527  LysGluAlaGluLysGluGluLysProGluValGluAsnAspLysGluAsnLeuLeu  546
          |||||||||
Db  1502  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  1561

Qy   547  LysGluLysThrAspAspThrSerGlyGluAspAspAsnAspAspGluAlaValAlaAlaSer  566
          ||||||||||||||||||||||||||||||||||
Db  1562  NNGGAGAAGACAGACGACACCTCAGGGGAGGACGACAACGACGAGGAAGAGGCCGTGGCCTCC  1621

Qy   567  LysGlyArgLysThrAlaAsnSerGlnIleValArgArgLysValArgIleThrArgSerMet  586
          ||||||||||||||||||||||||||||||||||||||
Db  1622  AAAGGCCGCAAAACTGCCAACAGCCAGGGAAGACGCAAAGGCCATCACCCGCTCAATG  1681

Qy   587  AlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnSerGlnLeuGluLeuAlaSer  606
          ||||||||||||||||||||||||||||||||||||
Db  1682  GCTAATGAGGCCAACAGCGAGGAGGCCATCACCCCCCAGAGCCAGCTGGCCTCC  1741

Qy   607  MetGluLeuAsnGluSerSerArgTrpProThrGluMetGlnTyrTrpAlaLysLysGly  626
          |||||||||||||||||||||||||||||||||||||
Db  1742  ATGGAGCTGAATGAGAGTTCTCGCTGGACAGAGAAGAAATGAAACAGCCAAGAAAAGT  1801

Qy   627  LeuLeuGluHisGlyArgAsnTrpSerArgAlaIleAlaArgMetValGlySerLysThrVal  646
          ||||||||||||||||||||||||||||||||||||||||
Db  1802  CTCCTGGACACGGCCGCAACTGGTCGGCCATCGCCGCCCGATGGTGGGCTCCAAGACTGTG  1861

Qy   647  SerGlnCysLysAsnPheTyrPheAsnTyrLysLysArgGlnAsnLeuAspGluIleLeu  666
          ||||||||||||||||||||||||||||||||||||||
Db  1862  TCGCAGTGTAAGAACTTCTATTCAACTACAAGAAGAGGCAGAACCTCGATGAGATCTTG  1921

Qy   667  GlnGlnHisLysLeuLeuMetGluLysMetGluLysArgAsnAsnIleArgArgLysLysAla  686
          ||||||||||||||||||||||||||||||||||||||||
Db  1922  CAGCAGCACAAGCTGAAGATGGAAGATGGAAGAAGGAAGAAGAGAAAGCG  1981

Qy   687  ProAlaAlaAlaSerGluGluLeuGluAlaAlaPheProProValGlnMetGluGlu  706
          |||||||||||||||||||||||||||||||||
Db  1982  CCGGCCGCGGCCAGCGAGGAGCTGCATTCCGCCCGTGTGTGGAGGATGGAG  2041
```

```
Qy    707 AlaSerGlyValSerGlyAsnGluGluMetValGluGluAlaGluAlaLeuHisAla 726
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   2042 GCGTCCGGCGTGAGCGGAAATGAGGAGGAGATGGTGGAGGAGGCTGAAGCNNNNNNN 2101

Qy    727 SerGlyAsnGluValProArgGlyGlyGluCysSerGlyProAlaThrValAsnAsnSerSer 746
          ||
Db   2102 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2161

Qy    747 AspThrGluSerIleProSerProHisThrGluAlaAlaAlaAspThrGlyThrAsnGly 766
          
Db   2162 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2221

Qy    767 ProLysProProAlaThrIleLeuGlyAlaAspGlyProProGlyProProThrProPro 786
          
Db   2222 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2281

Qy    787 ArgArgThrSerArgAlaProIleGluGluProThrProAlaSerGluAlaThrGlyAlaPro 806
          
Db   2282 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2341

Qy    807 ThrProProAlaProProSerProSerAlaProProProValProLysProGluGlu 826
          
Db   2342 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2401

Qy    827 LysGluGluGluThrAlaAlaAlaAlaAlaProProValGluGlyGluGlyGlnLysProProPro 846
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2402 NAGGAGGAGAACCGGCAGCGCCCCCCAGTGGAGGGGGAGGAGCAGAAGCCCCCC 2461

Qy    847 AlaAlaGluGluLeuValAlaValAspThrLysThrGlyLysLysAlaGluGluProValLysSerGluCys 866
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2462 GCGNNNGAGGAGCTGGCAGTGGCAGTGGACACAGGGAAGGCTGGAGCCCGTCAAGAGCGAGTGC 2521

Qy    867 ThrGluGlyAlaGluGluGlyProAlaLysGlyLysAspAlaAlaAlaGluAlaAlaThr 886
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2522 ACGGAGGGGGCCGAGGAGGGCCCGGCCAAGGACGCGCCGAGGGCGTACG 2581

Qy    887 AlaGlyGluLysAlaAlaLeuLysLysAlaGluGluLuLysAlaLysGlyLysGlyGlySerGlyGlyArgGlyAlaThrThrAla 906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2582 GCCGGAGGGGCGCTCAAGCAGCAGAAGAAGAAGAGGGCGGAGGACAGCGGCGGCGGCCCACCACAGCC 2641

Qy    907 LysSerSerGlyAlaProGlnAsnAspSerGlnSerSerSerAlaThrCysSerSerArgAlaAspGluVal 926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2642 AAGAGCTCGGGCGCCCCCGGAGAGCAGCGACTCCAGTGCCACCTGCAGTCGAGACGAGGTG 2701

Qy    927 AspGlyAlaGluGluGlyAlaAsnAlaSerAsnArgLeuLeuSerProArgProSerLeuLeuThr 946
          |||||||||||||||||||||||||||||||||||||||||||||
Db   2702 CCGACTGGCGCAGAACCGGCTCGTTGCTGTCCCCAGAGGCCTCCTCACC 2761

Qy    947 AspGluAlaGluGlyGlyAspAspLysProLysAsnAlaSerProGlnLysProLeuAspLeuLysProLeuGlnLeu 966
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2762 GATGAGGCCGAGGGCGGCGACAAGCCGAAGAACGCCTCCCCACACTGCAGCAGCACGCTG 2821

Qy    967 LysGlnArgAlaAlaAlaAlaAlaProAlaAlaIleLeuProProGlnLysThrLeuLysLysGluProProArg 986
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2822 AAGCAGCGAGCGGCCGCCGCCGCCCCGGCAGCCATCCTGCCACCCCAGAAAACCCTGAAG 2881

Qy    987 GluAspProAlaAlaProAlaGlnGlnArgGlyGlnProGlnLysGlySerSerProArgPro 1006
          |||||||||||||||||||||||||||||||||||||||||||||
Db   2882 GAGGACCCGGCAGCTCCCGCCCAGCAGCGCGGGCCAGCCCCAGAAAGGCAGCAGCCCCAGGCCG 2941

Qy   1007 ProGluSerAspProGlnGlnGlnGlnProProProAlaLysGlnLysProGlyGlyArgGlyAspPro 1026
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2942 CCGGAGAGCGACCCGCAGCAGCCTGCAGCAGCCTCCGCCAAGCAGAAGCCAGGAGGCCGCGGGGACCCG 3001

Qy   1027 AlaAlaProAlaAlaAspLysLysAlaAlaPheHisAlaAlaAlaGlnLysLysProGlyLeuGlyAspPro 1046
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3002 GCACCCCCGCCCCACAAGAAGGCGGCCTTCCATGCCGCGGCCCAGAAGAAGCCAGGGCTCGGAGACCCG 3061

Qy   1047 ProCysTrpThrSerGlyLeuProPheProValProProArgGluValIleLeuLysAlaSer 1066
          |||||||||||||||||||||||||||||||||
Db   3062 CCATGCTGGACCTCGGGCCTGCCCTTCCCAGTGCCACCTCGTGAAGTCATCCTCAAGGCCTCC 3121

Qy   1067 ProHisAlaProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeu 1086
          ||||||||||||||||||
Db   3122 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTCACCCAGTGCCCCTG 3181

Qy   1087 GlyLeuHisAspThrAlaArgProValLeuAspProProThrIleSerAsnProPro 1106
          ||||||||||||||||||||||||||||||||||||||||||||
Db   3182 GGCTTCCATGACACTGCCCGGCCGTCCTGCCGCGGCCCACCACCATCTCCAACCCGCCT 3241

Qy   1107 ProLeuIleSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyGlyAlaIleSer 1126
          |||||||||||||||||||||||||||||||||||||||||||||
Db   3242 CCCCTTATCTCCTCTGCCAAGCACCCCAGCGTCCTCGAGAGGCAAATAGGTGCCATCTCC 3301

Qy   1127 GlnGlyMetSerValGlnLeuHisValProTyrSerGluHisAlaAlaLysAlaProValGly 1146
          |||||||||||||||||||||||||||||||||||||||||||||
Db   3302 CAAGGAATGTCGGTGCAGCTCCACGTCCCGTACTCAGAGCATGCCGCAGGCCCGGTGGGC 3361

Qy   1147 ProValThrMetGlyLeuProLeuProMetAspProLysLeuLeuAlaProPheSerGly 1166
          |||||||||||||||||||||||||||||||||||||||||
Db   3362 CCCGTCACCATGGGGCTGCCCCTGCCCATGGACCCCAAAAAGTCGGCACCCTTTNNGCGGA 3421

Qy   1167 ValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProGluSerGluGlyValGlyVal 1186
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   3422 GTGAAGCAGGAGCAGCTGTCCCCACGGGGCCAGGCTGGGCACCGGAGAGCCTGGGAGTG 3481

Qy   1187 ProThrAlaGlnGluAlaSerValLeuValArgGlyThrAlaLeuGlySerValProGlyGly 1206
          |||||||||||||||||||||||||||||||||||||||||||
Db   3482 CCCACGGCCCAGGAGGCATCCGTGCTGGTGAGGGGCACAGCTCTGGGCTCAGTTCCGGCGGA 3541

Qy   1207 SerIleThrLysGlyIleProSerThrArgValProSerArgAspSerAlaIleThrTyrArg 1226
          |||||||||||||||||||||||||||||||||||||||||||||
Db   3542 AGCATCACCAAAGGCATTCCCAGCACACGGGTGCCCTCGGACAGCGCCATCACATACCGC 3601

Qy   1227 GlySerIleThrHisGlyThrProAlaAspValLeuTyrTyrLysGlyThrIleThrArgIle 1246
          |||||||||||||||||||||||||||||||||||||||||||
Db   3602 GGCTCCATCACCCACNNNACGCCAGCTGACGTCCTGTANNNNGGCACCATCACCAGGATC 3661

Qy   1247 IleGlyGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeuProLysGlySly 1266
          |||||||||||||||||||||||||||||||||||||||||||
Db   3662 ATCGGCGAGGACAGCCCNNGTCGCTTGGANNGCGGCCGGAGGACAGCCTGCCCAAGGGC 3721

Qy   1267 HisValIleTyrGluGlyLysLysGlyHisValValLeuSerTyrGluValLeuGlyMetSerVal 1286
          ||||||||||||||||||||||||||||||||||||||||||
Db   3722 CACGTCATCTTACGAGGCAAGAAGGGCCACGTCGTTCTTATGAGANNNNNNNNNNN 3781

Qy   1287 ThrGlnCysSerLysGluAspGlyArgSerSerGlyProHisGluThrAlaAla 1306
          
Db   3782 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3841

Qy   1307 ProLysArgThrTyrAspMetMetGluGlyGluValGlyArgValGlyArgAlaIleSerSerAlaSer 1326
          
Db   3842 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3901

Qy   1327 IleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisLeuLys 1346
          
Db   3902 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3961

Qy   1347 GlyGlnHisHisIleArgGlySerIleThrGlnIleProArgSerTyrValGluAla 1366
          
Db   3962 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4021

Qy   1367 GlnGluAspTyrLeuArgArgGluArgLeuLeuLysArgGluGlyThrProProPro 1386
          
Db   4022 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4081

Qy   1387 ProProProSerArgAspLeuThrGlnAlaTyrLysThrGlnAlaLeuGlyProLeuLys 1406
          
Db   4082 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4141

Qy   1407 LeuLysProAlaHisGluGlyLeuValAlaAlaThrValLysGluAlaGlyArgSerIleHis 1426
          
Db   4142 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4201

Qy   1427 GluIleProArgGluGluLeuArgHisThrProGluLeuProLeuAlaProArgProLeu 1446
```

Db 5282 AAACCAACCACCACATCCTCGTCCGAGCGGAGCGGAGGAGACCGGATCGAGAGCGCNNNNNN 5341
Qy 1807 AspArgGluLysSerIleLeuThrSerThrThrValGluHisAlaProIle 1826
Db 5342 NNNNNNGAGCGGGAAAAGNCCATCTCACGTCCACNNNNGNNNGTGGAGCACGCCCCATC 5401
Qy 1827 TrpArgProGlyThrGluGlnSerSerGlySerSerGlyGlyGlyGlyGly 1846
Db 5402 TGGAGNNNNNGTACAGAGCAGAGGCAGCAGCGGCAGCAGCGGTGGGGGTGGGGGC 5461
Db 1847 SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProArg 1866
Db 5462 AGCAGCAGCGCGCCGCCTCCCACTCCCANNCCACCAGCACTCGCCCATCTCCCCTNNN 5521
Qy 1867 ThrGlnAspAlaLeuGlnLeuGlnArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1886
Db 5522 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5581
Qy 1887 IleThrAlaValGluProSerLysProThrValArgSerThrSerSerPro 1906
Db 5582 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5641
Qy 1907 ValArgProAlaAlaThrProPheProProAlaThrHisCysProLeuGlyThrLeuAsp 1926
Db 5642 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5701
Qy 1927 GlyValTyrProThrLeuMetGluProValLeuLeuProLysGluLeuProLysAlaProArgValAla 1946
Db 5702 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5761
Qy 1947 ArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProProAlaArg 1966
Db 5762 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5821
Db 1967 SerGlyLeuGluProAlaSerSerProSerLysSerGlyProProGluProLeuValPro 1986
Db 5822 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGGCCCCTNGTGCCC 5881
Qy 1987 ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHis 2006
Db 5882 CCCGTCTCTGGCCACGCGCCACCATCGCCCGCCACCCTGCGAAGAACCTCGCACCTCACCAC 5941
Qy 2007 AlaSerProAspProProAlaProAlaSerArgAlaSerProHisArgGluLysThr 2026
Db 5942 GCCAGCCCCGACCCCGCCCCACCTGCCTCCGCTCCGGACCCGCACCGGGAAAAGACT 6001
Qy 2027 GlnSerLysProPheSerIleGlnLeuGlnLeuGlnLeuArgGlnLeuSerGlySer 2046
Db 6002 CAAAGTAAACCCTTTTCCATCCAGGAACTGGAACTCCGTTCTCTGGNNNNNNNNNNN 6061
Qy 2047 SerTyrSerProGluGlyValGluProValSerProSerProSerLeuThrHis 2066
Db 6062 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6121
Qy 2067 AspLysGlyLeuProLysHisGluGluLeuAspLysSerHisLeuGluGluLeu 2086
Db 6122 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6181
Qy 2087 ArgProLysGlnProValLeuGlyValGlyGluAlaAlaHisLeuProHisLeu 2106
Db 6182 NNNNNNNNNNNNNNNNNGCCCGTGAAGTTGGCGGGGAGGCCGCCACCTGCACCACCTG 6241
Qy 2107 ArgProLeuProGlnSerGlnProSerSerProLeuGlnThrAlaProGlyVal 2126
Db 6242 CGGCCCGCTGCCTGAGCAGCCAGCCCGTGCTCCAGACCGCCCCAGGGGGTC 6301
Qy 2127 LysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAsp 2146
Db 6302 AAAGGTCACCAGCCGGGTGGTCACCCTGGCCNNNNNNNNNNNNNNNNNNNNNNNNN 6361
Qy 2147 TyrThrArgHisHisProGlnGlnLeuSerAlaProLeuProLeuTyrSerPhe 2166
Db 6362 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6421

Db 4202 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4261
Qy 1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
Db 4262 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4321
Qy 1467 GlySerLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProPro 1486
Db 4322 NNNNNNNNNNACGTACGTCCCTCATCGGCAGCCCGGGACGTTCCCACCC 4381
Qy 1487 ValHisProLeuAspValMetAlaAspAlaArgAlaLeuLeuGluArgAlaCysTyrGluGlu 1506
Db 4382 GTGCACCCGCTGGATGTGATGGCCGACGCCCGGGCACTGGACGTGCCTGCTACAGGAGAG 4441
Qy 1507 SerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAlaAlaArgGlyAla 1526
Db 4442 AGCCTGAAGAGCCGGCCAGGCNNNNNNNNNAGCTCGGGGGCTCCATTGCGCGCGGCGCC 4501
Qy 1527 ProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHis 1546
Db 4502 CCGGTCATTGTGCCTGACCTGGGCAAGCCGCGGCAGAGCCCCCTGACCTATGAGGACCAC 4561
Qy 1547 GlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluProThr 1566
Db 4562 GGGGCACCCTTTGCGGGCCACCTCCCCGAGGTTCGCCCGTGACCACGCGGGAGCCCACG 4621
Qy 1567 ProArgLeuGlnGluGlyLeuSerLeuSerSerLysSerAlaSerGlnAspArgLysLeuThr 1586
Db 4622 CCGCGCCTGCAGGAGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4681
Qy 1587 SerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisProHis 1606
Db 4682 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4741
Qy 1607 ProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSer 1626
Db 4742 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4801
Qy 1627 HisIleProLeuAlaPheAspProThrIleProLeuProArgGlyIleProLeuAspAlaAla 1646
Db 4802 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCC 4861
Qy 1647 AlaAlaTyrTyrLeuProArgHisLeuProAlaProAsnProThrTyrProHisLeuTyrPro 1666
Db 4862 GCTGCCTACTACCTGCCCCGGCACCTGCCCGCCCCCAACCCCACCTACCCCA 4921
Qy 1667 ProTyrTyrLeuProArgGlyTyrProAspThrAlaLeuGluAlaAsnArgGlnThrIleIle 1686
Db 4922 CCCTACCTCATCCGCGGCTACCCCGACACAGCGCTGGAGAACCGGCAGACCATCATC 4981
Qy 1687 AsnAspTyrIleThrSerGlnGlnMetHisHisAsnProHisAlaThrAlaMetAlaGlnArg 1706
Db 4982 AATGACTACCATCACCTCGCAGCAGATGCACCACAACGCGGCCACCATGGCCCAGCGA 5041
Qy 1707 AlaAspMetLeuArgGlyLeuSerProArgGluSerProLeuAlaLeuAsnTyrAlaAla 1726
Db 5042 GCTGACATGCTGAGGGGCCTTCTCGCCCCGAGTCCTCGGCACTCAACTACGCTGCG 5101
Qy 1727 GlyProArgGlyIleIleIleAspLeuSerGlnValProHisLeuProValLeuProPro 1746
Db 5102 GGTCCCCGAGNNNNNNNNNNNNNNGCCCGTGAAGTTGGCGGGGAGGCCGCCACCTG 5161
Qy 1747 ThrProGlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGln 1766
Db 5162 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5221
Qy 1767 ProPheSerArgHisSerSerSerProLeuSerProGlyGlyProThrHisLeuThr 1786
Db 5222 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5281
Qy 1787 LysProThrThrSerSerArgGluArgAspAspArgArgGluArgAspArg 1806

```
source          1..2000
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5016291"
                /tissue_type="Lung, small cell carcinoma"
                /clone_lib="NIH_MGC_7"
                /lab_host="DH10B-R"
                /note="Vector: pOTB7"

ORIGIN

Alignment Scores:
Pred. No.:                          Length:         2000
Score:          1.29e-96            Matches:        503
Percent Similarity:   90.72%        Conservative:   15
Best Local Similarity: 88.09%       Mismatches:     31
Query Match:    19.42%              Indels:         22
DB:             11                  Gaps:           1

US-09-522-753-5 (1-2517) x BC033087 (1-2000)

Qy    1  MetSerGlySerThrGlnLeuValAlaGlnThrTrpArgAlaThrArgAlaThrGluProArgTyrPro   20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  273  ATGTCGGATCCACCAGCCTGTGGCACAGAGCTCGCACAGAGCTGGAGGGCCACTGAGCCCGCTACCCG  332

Qy   21  ProHisSerLeuSerTyrProValGlnIleAlaArgThrArgHisThrAspThrArgValGlyLeuLeu   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  333  CCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGACGCGACACACGGACGTCGGGGCTCCTG  392

Qy   41  GluTyrGlnHisHisSerArgArgAspTyrAlaSerHisSerLeuSerProGlySerIleIleGln   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  393  GAGTACCAGCACCACTCCCGGACTATGCCCACCGTTGCCCACCGTCGCCCGGCTCCATCATCCAG  452

Qy   61  ProGlnArgArgProSerLeuLeuSerGluProGlnPheGlnAsnGluArgSerGln   80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  453  CCCCAGCGGCGGCCGAGCCCCTCCCTGTCTGTCGAGTTCCAGCCCCGGAATGAACGGTCCCAG  512

Qy   81  GluLeuHisLeuArgProGluSerHisSerThrLeuProGluGlyLysSerGluMet  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  513  GAGCTCCACCTGCGCCCAGAGTCCCACTCCACCCTGCCCGAGGGCTGGGGAAGCTCAGAGATG  572

Qy  101  GluPheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgPro  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  573  GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTAGAGCTGCTGCCTGCTGCGACCG  632

Qy  121  SerProLeuLeuAlaThrGlyGlnAsnProAlaGlySerGluThrLysAspArgSer  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  633  TCACCCCTGCTGGCCACGGGCCAGCGCGCGGATCGAAGACCTTGAAGACCTCCACCAAGGACCGTAGC  692

Qy  141  LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu  160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  693  CTGACGGGCAAGCTGGAACCGGTGTCTCCCCCAAGCCCCCCGCACACTGACCCTGAGCTG  752

Qy  161  GluLeuValProProArgLeuSerLysSerGluLeuIleGlnAsnMetAspArgValAsp  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  753  GAGCTGGTGCCGCCACGGCTGTCCAAGAGGAGAGTGATCCAGAACATGGACCGCGTGGAC  812

Qy  181  ArgGluIleThrMetValGluGlnGlnIleLeuSerLysLeuLysGlnGlnLeuLeu  200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  813  CGGGAGATCACCATGGTAGAGCAGCAGATCTCAAGCTGAAGAAGAGCAGCAACAGCTG  872

Qy  201  GluGluAlaAlaLysProGluProProGluLysSerProValSerProProProIleGlu  220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  873  GAGGAGGAGGCTGCCAAGCCGCGGCCCGTGTCACCGCCCCCATCGAG  932

Qy  221  SerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnArgLysAlaGluAla  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  933  TCGAAGCACCGCAGCCTGGTGCAGATCATCTACGACGAGAACCGGAAGAGGCTGAAGCT  992

Qy  241  AlaHisArgIleLeuGluGlyLeuGlyTyrProGlnValGluLeuProLeuTyrAsnGln  260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  993  GCCATCAGCGGATTCTGGAAGCCCTGGGGCCCCCAGGTGGAGCTGCCGCTGTACAACCAGCCC  1052
```

```
Qy 2167  ProGlyAlaSerCysProValLeuArgArgProAspLeuArgTyrLeuPro  2186
Db 6422  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  6481

Qy 2187  ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSer  2206
Db 6482  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTCT  6541

Qy 2207  ProGluProAsnLysThrValGlyGlyGlyGluAspGlyIleGluProValSer  2226
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6542  CCAGAGCCAAACAGACGTCAGTCTTGGGTGGCGGTGAGGACGGTATTGAACCTGTGTCC  6601

Qy 2227  ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr  2246
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6602  CCACCGGAGGGGGCATGAGCGGAGCCAGGGCACTCCCGAGTGCTGTGTACCCGCTGCTGTAC  6661

Qy 2247  ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer  2266
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6662  CGGGATGGGGAACAGACAGACCGAGGATGGGCTCCAAGTCTCCAGGCAACACCAGC  6721

Qy 2267  GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys  2286
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6722  CAGCCGCCCAGCCCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAG  6781

Qy 2287  LysGlnGluIleLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluThrTyr  2304
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 6782  AAGCAAGAGATCAACAAGAAGCTGAACACCACAACCCACAACCGGAATGAGCCTGAATAC  6835
```

```
RESULT 4
BC033087
LOCUS       BC033087               2000 bp    mRNA    linear    HTC 04-MAR-2003
DEFINITION  Homo sapiens, Similar to nuclear receptor co-repressor 2, clone
            IMAGE:5016291, mRNA.
ACCESSION   BC033087
VERSION     BC033087.1    GI:21619972
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 43 Row: i Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 5454073
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
```

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
COMMENT   Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm .tmc.edu.
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M.,Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R.,
            Lawrence, S., Richards, S., Gibbs,R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 12 Row: i Column: 23
            This clone has the following problem: no polyA-tail.
FEATURES   Location/Qualifiers
     source   1..1362
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4179307"
                     /tissue_type="Brain, anaplastic oligodendroglioma with
                     1p/19q loss"
                     /clone_lib="NCI_CGAP_Brn67"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"

ORIGIN
Alignment Scores:
Pred. No.:   2.7e-84   Length: 1362
Score: 2264.00   Matches: 431
Percent Similarity: 99.54%   Conservative: 0
Best Local Similarity: 99.54%   Mismatches: 2
Query Match: 17.13%   Indels: 0
DB: 11   Gaps: 0

US-09-522-753-5 (1-2517) x BC020427 (1-1362)

Qy   1676 ThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnMet 1695
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ACGGCGGCGCTGGAGAACCGGCAGACCATCATCAATGACTACATCACCTCGCAGCAGATG 122

Qy   1696 HisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerPro 1715
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 CACCACAACGCGGCCACCGCCATGGCCCAGCGCGCTGATATGCTGAGGGGCCTCTCGCCC 182

Qy   1716 ArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyIleIleAspLeuSer 1735
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 CGCGAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCCGGCATCATCGACCTGTCC 242

Qy   1736 GlnValProHisLeuProValLeuValProThrProGlyThrProAlaThrAlaMet 1755
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CAAGTGCCACACCTGCCTGTGCTCGTGCCCACACCAGGCACCCCAGCCACCGCCATG 302

Qy   1756 AspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerArgHisSerSerSer 1775
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 GACCGGCTGGCCTTGCCTACCTCCCCACCGCCCCAGCCCTTCAGCAGGCACAGCAGCTCC 362

Qy   1776 ProLeuSerProGlyGlyProThrHisLeuThrLysProThrThrThrSerSerSerGlu 1795

RESULT 5
BC020427
LOCUS   BC020427   1362 bp   mRNA   linear   HTC 03-JAN-2002
DEFINITION Homo sapiens, clone IMAGE:4179307, mRNA.
ACCESSION   BC020427
VERSION   BC020427.1 GI:18043012
KEYWORDS   HTC.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens

Qy   261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCAGATGCGGAAGAAGCTA 1112

Qy   281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
          ||||||||||||||||| ||||||||||||| :::||||||||||||||||
Db   1113 ATCTTGTACTTCAAGAGGAGAAACCATCACGCTCGGAAACAATGGGAGCAGAAGTTCTGCCAG 1172

Qy   301 ArgTyrArgAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsnAsnPro 320
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 CGCTATGACCAGCTCATGGAGGCCTGGAGAAGAAGGTGGAGCGCATCGAGAACAACCCCC 1232

Qy   321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 CGGCGGCGGGCCAAGGAGCAAGGTGCGCGAGTACTACGAGAAGCAGTTCCCTGAGATC 1292

Qy   341 ArgLysGlnArgGluLysGlnLeuGluArgMetGlnSerArgValGlyGlnArgGlySerGly 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 CGCAAGCAGCGGGAGCTGCAGAGCAGACAGGAGAGCAGGGTGGGCCAGCGAGGCAGTGGG 1352

Qy   361 LeuSerMetSerArgAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CTGTCCATGTCGCGCGCCGCCCGGAGCGAGCACGAGGTGTCAGAGATCATCGATGGCCTCTCA 1412

Qy   381 GluGlnGluAsnLeuGluLysGlnMetArgGlnMetArgGlnGlnMetGlnAlaProProMetTyr 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GAGCAGGAGAACTTGGAGAAGCAGATGCGCCAGCTGCCCGTGATCCCCATGCTGTAC 1472

Qy   401 AspAlaAspGlnGlnAsnGlnIleLeuLysPheIleLysAsnMetAsnGlyLeuMetAlaAspProMet 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GACGCTGACCAGCAGCGCCATCAAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATG 1532

Qy   421 LysValTyrLysAspArgArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPhe 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533 AAGGTGTACAAAGACCGGCCAGGTCATGAACATGTGGAGTGAGCAGGAGAAGAGAGACCTTC 1592

Qy   441 ArgGluGluLysPheMetGlnHisProLysAsnProGlyLeuIleAlaSerPheLeuGluArg 460
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1593 CGGGAGGAAGTTCATGCAGCATCCCAAGAACTTGGCCTGATCGCATCATTCCTGGAGAGG 1652

Qy   461 LysThrValAlaGluCysValLeuTyrTyrTyrLeuThrTyrTyrLeuAsnGluAsnTyrLys 480
          |||||||||||||||||||||||| :::::||||||
Db   1653 AAGACAGTGGCTGAGTGCGTCCTCTATTACTACCTGACTTAAGAAGAAATGAGAACTATAAG 1712

Qy   481 SerLeuValArgArgSerTyrArgArgArgGlyArgGlyTyrGlyLysSerGlnGlnGlnGlnGlnGln 500
          ||||||||||||||||||||||| ::: ::
Db   1713 AGCCTGGTGAGACGGAGCTATCGGCGGCCGGCTGCTCCAAAAGTCTCAGAGGGTCAAAAAGTCCAGCCC 1772

Qy   501 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProArgSerGlnArgGlyPheSerGlnSerGlnGluGluLys 520
          :::::::
Db   1773 CAAAAAAGATCACCGGCCGGCCGGAGTAAAAAAGGCTTCCAGCCCGAGAAGGTTCCTGCCCAGAAAG 1832

Db   521 AspGluLysGluLysLysGluGluAlaGluLysGlyGluGluLysLysProGluValGluAsn 540
          ||||||||||||||||||||||||| :::
Db   1833 CCACAGGCCGAGAAAGCAGCGCGCCTGCTCAAAAGTTCAGAGGGTCAAAAAAGTCCAGCCCC 1892

Qy   541 AspLysGlu----------- 543
          |||||
Db   1893 AGAAAGCACCTGCTCCAAAAGCATCTGCAAGAAAAGCATAAGTGGCAATCATCAAAAAAGTA 1952

Qy   544 -----------AspLeuLeuLysGluLys 549
          ::||:::::||
Db   1953 ATAAAAGGTTCTTTTTGACCTGTTAAAAAAAAA 1985

```
VERSION     BC026028.1  GI:19684000
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1966)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 42 Row: b Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 5454137
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..1966
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3836580"
                     /tissue_type="Ovary, adenocarcinoma"
                     /clone_lib="NIH_MGC_9"
                     /lab_host="DH10B-R"
                     /note="vector: pOTB7"
ORIGIN
Alignment Scores:
Pred. No.:                                          Length:              1966
Score:           1.99e-58                           Matches:              343
Percent Similarity:  73.35%                         Conservative:          56
Best Local Similarity: 63.05%                       Mismatches:            98
Query Match:     12.44%                             Indels:                47
DB:              11                                 Gaps:                  10

US-09-522-753-5 (1-2517) x BC026028 (1-1966)
```

```
Qy     16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
          ||||||||||||||||||||||||   :::::                |||||||||||
Db    392 CAAAGTCGTTATCCTCCTCACTCTGTCCAGTATACATTTCCAACACCCGCCACCAGCAG 451

Qy     36 AspValGlyLeuLeuGluTyrGln------HisHisSerArgAspTyrAlaSerHisLeu 53
          :::  :::::|||:::                   |||||||||||||||||||||||||
Db    452 GAGTTCGGACTCCTGGATTATCGTTCCTCCTCATCTTGAAGTGAGTCAGGCATCACAGCTT 511

Qy     54 SerProGlySerIleIleGlnProGlnArgArgProSerLeuSerGluGluPheGln 73
          |||||||||:::|||     |||||||||||||||||||||||||||||||||
Db    512 TTGCAGCAACAGCAGCAGCAACAGCTTCGAAGGCGACCTTCCTTGCTTTCAGAATTCAC 571

Qy     74 ProGlyAsnGluArgSerGlnGluHisLeuArgProGluSerHisSerTyrLeuPro 93
          ||||||:::|||  |||     :::|||
```

```
Db    363 CCACTCTCCCAGGAGGTCCAACACACTTGACAAAACCAACCACCACGTCCTCGTCCGAG 422
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1796 ArgGluArgAspArgAspArgAspArgAspArgGluArgGluLysGluSerIleLeu 1815

Db    423 CGGGAGCGGAGACCGGGATCGAGAGCGGGACCGGGACCGGGAGCGGGAAAAGTCATCCTC 482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1816 ThrSerThrThrValGluHisAlaProIleTrpArgProGlyThrGluGlnSerSer 1835

Db    483 ACGTCCACCACGACGGTGGAGCACGCACCCATCTGGAGACCTGGTACAGACAGAGCAGC 542
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1836 GlySerSerGlySerGlyGlyGlyGlySerSerSerArgProAlaSerHisSer 1855

Db    543 GGCAGCAGCGGCAGCGGGCGGGTGGGGCAGCAGCGGCAGCGGCCCGGCCTCCCACTCC 602
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1856 HisAlaHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgPro 1875

Db    603 CATGCCCACCAGCACTGCCCCATCTCCCCTCGGACCCAGGATGCCCTTCCAGCAGAGACCC 662
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1876 SerValLeuHisAsnThrGlyMetLysGlyValIleIleIleThrAlaValGluProSerLysPro 1895

Db    663 AGTGTGCTTCACAACACAGGCATGAAGGGTATCATCACCGCTGTGGAGCCCAGCACGCCC 722
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1896 ThrValLeuArgSerThrSerProValArgProAlaAlaThrPhePheProPro 1915

Db    723 ACGGTCCTGAGGTCCACCTCCACCTCCACCGTTCGCCCCGGCTGCCACATTCCCACCT 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1916 AlaThrHisCysProLeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluPro 1935

Db    783 GCCACCCACTGCCCACTGGGCGGCACCCTCGATGGGGTCTACCCTACCCTCATGGAGCCC 842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1936 ValLeuLeuProLysProLysGluAlaProAlaArgValAlaArgProGluProArgAlaAspThr 1955

Db    843 GTCTTGCTGCCCAAGCAGAGGAGCCCCGGGTCGCCCCAGAGCGGCCCCGAGCAGACACC 902
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1956 GlyHisAlaPheLeuAlaLysProProAlaAlaArgSerGlyLeuGluProAlaSerSerPro 1975

Db    903 GGCCATGCCTTCCTGCCAAGCCCCCAGCCCGCTGGGCCGCTGAGCCCGCCTCCTCCCCC 962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1976 SerLysSerGlySerGluProArgProLeuValProProValSerGlyHisAlaThrIleAla 1995

Db    963 AGCAAGGGCTCGGAGCCCCGGCCCCTCGTGCCTCCCGTGTCTGGCCACGCCACCATCGCC 1022
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1996 ArgThrProAlaLysAsnLeuAlaProHisHisAlaSerProAspProAlaProPro 2015

Db   1023 CGCACCCCTGCGAAGAACCTCGCACCTCACCACGCCAGCCCGGACCCGGCGCCACCT 1082
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2016 AlaSerAlaSerAspProHisAlaArgGluLysThrGlnSerLysSerProPheSerIleGluGlu 2035

Db   1083 GCCTCCGGCCCTCCGGACCCCACCGGGAAAAAGACTCAAAGTAAACCCTTTCCATCCAGGAA 1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2036 LeuGluLeuArgSerLeuGlyTyrHisGlySerSerTyrSerProGluGlyValGluPro 2055

Db   1143 CTTGGAACTCCGTTCTCTGGGTTACCACGGCAGCAGCTACAGCCCCGAGAGGGTGGAGCCC 1202
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2056 ValSerProValSerProSerLeuThrHisAspLysLysProLysHisLeuGlu 2075

Db   1203 GTCAGCCCCGTGAGTCCCAGCCTCACCCCACGACAAGGGGCTCCCCAAGCACCTGAA 1262
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2076 GluLeuAspLysSerHisLeuGluGlyLeuGlyLeuArgProLysGlnProGlyValLys 2095

Db   1263 GAGCTGGACAAGAGCCACCTGGAGGGGGAGCTTGGCCTCAGGCCGAAGCAGCCCGGAGTGAAG 1322
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   2096 LeuGlyGlyGluAlaAlaHisLeuProHisLeuArgPro 2108
          |||||||||||||||||||||||||||||||||||||||
Db   1323 CTTGGGGGGGAGGCCGCGCCCACCTCCCACACCTGCGGCCG 1361
```

```
RESULT 6
BC026028
LOCUS       BC026028        1966 bp    mRNA    linear    HTC 22-MAR-2002
DEFINITION  Homo sapiens, Similar to nuclear receptor co-repressor 1, clone
            IMAGE:3836580, mRNA.
ACCESSION   BC026028
```

```
Db   572  CCAGGTTCTGCAGAGGCCTCAAGA------AGGAGAACTAGTTATGACCGTTCATCCA  625
                ||||||    :::                  |||||||||||||||||||||
Qy    94  GluLeuGlyLysSerGluMetGlyThrPheIleGluSerLysArgProArgLeuGluLeu  113

Db   626  GGCCCATCCCCAGTGGATCATGATCATGATTCACTCGAATCGAAGCGACCACGTCTCGGAACAGGTT  685
                :::                                                :::
Qy   114  ProAsp--------------------ProLeuLeuArgProSerPro  122

Db   686  TCTGATTCTCATTTTCAGCGTGTCAGTGCTCGCGGTTTGCCTTTAGTGCACCCGCTGCCA  745
                |||||||    :::                ||   |||
Qy   123  LeuLeuAlaThrGlyGlnProAlaGlyLysSerGluAspLeuThrLysArgArgSerLeuThr  142

Db   746  -------GAAGGGCTGAGGGCT--TCTGCAGATGCTAAGAAGGATCCAGCATTCGGA  793
                                |||       :::
Qy   143  GlyLeuGlyLeu---ProValSerProSerProProHisThrAspProGluLeuGlu  161

Db   794  GGCAAACATGAAGCTCCATCCTCCAATTTCGGGGCAACCATGTGGAGATGATCAAAAT  853
                ||||                          :::
Qy   162  LeuValProProArgLeuSerLysGluGluLysGlnAsnMetAspArgValAspArg  181

Db   854  GCTTCACCTTCAAAACTTCAAAGGAAGGAGAGTTAATACAGAGTATGGATCGTGTAGATCGA  913
                |||||
Qy   182  GluIleThrMetValGluGluGlnArgIleSerLeuLysLysGlnGlnGlnGlnLeuGlu  201

Db   914  GAAATTGCAAAAGTAGAACACAGACAGATCCTTAAACTGAAAAAGAAACAACACACAGCTTGAA  973
                ||||||||||||
Qy   202  GluGluGlnAlaAlaLysProGluProGluLysSerProValSerProProIleGlnSer  221

Db   974  GAAGAGGCAGCTAAACCTCCTGAGCCTGAGAAGCCCGTGTCCCCTCCTCGTGGAGCAG  1033
                ||||||||||||
Qy   222  LysHisArgSerLeuValGlnIleIleIleTyrAspGluAsnArgLysLysAlaGluAlaAla  241

Db  1034  AAACACCGACTATTGTCCAAATTATTTATGATGAGAATCAGAAAAGACAGAAGAAGCT  1093
                ||:::||||||
Qy   242  HisArgIleLeuGluGlyLyrProGlnValGlyLeuProLeuTyrAsnGlnProSer  261

Db  1094  CATAAAATTTTTGAAGGTCTTGGCCCAAAGTTGAACTGCCACTGTATAACCAGCCATCA  1153
                |||||||||
Qy   262  AspThrArgGlnTyrHisGluAsnIleLysSerIle------  272

Db  1154  GATACCAAGGTGTACCATGAACATCAAGACTGGAGTACCTGCAAGGCGCATGATGAAA  1213
                |||||||
Qy   273  AsnGlnAlaMetArgLysLysLeuLeuLeuTyrPheLysArgArgAsnHisAlaArgLys  292

Db  1214  AACCAGTGATGAGGAAAAACTCATTTTATTTTTTAAAAGAAGAAAATCATGCAAGAAA  1273
                ||||||
Qy   293  GlnTrpLysGlnLysPheCysGlnArgTyrAspProGlnLeuMetGluAlaLeuGluLysLys  312

Db  1274  CAAAGGGACAAAAAATCTGCCAGCGTTATGATCAGCTCAGTGGAGGCATGGGAGAAAAAA  1333
                :::
Qy   313  ValGluArgIleGluAsnAsnProArgArgArgAlaLysAlaArgGluSerLysValArgGluTyr  332

Db  1334  GTTGGACAGAATAGAAAATAATCCTCGGAGAAGAAAGCTAAAGAAACAAGGGAATAC  1393
                |||||
Qy   333  TyrGluLysGlnPheGlnLeuLeuArgLysArgLysGlnGluLeuGlnLeuGlnArgMetGlnSer  352

Db  1394  TATGAAAAGCAGTTTCAGAAATTCAAAAATTCGAAAACAAGAGAAACAGCAAGAAAAGATTTCAG---  1450
                ||||||||             ||||||
Qy   353  ArgValGlyGlnArgGlyLeuSerGlyMetSerSerAlaAlaArgSerGluHisGluVal  372

Db  1451  CGAGTTGGGCAGCAGGGGAGCTGGTCTTGTCTTCAGCCACCATTGCTAGGAGTGAGCATGAGATT  1510
                |||||||
Qy   373  SerGluIleIleAspGlyLeuAsnLeuSerGlyGluGlnLysLeuAsnLeuGluLysArgGlnAsnLeu  392

Db  1511  TCTGAAATTATTGATGGGCTCTCTGAGCAGGAGCAGAAGAAATGAGAAACAAATGCGGCAGCTC  1570
                ||||||
Qy   393  AlaValIleProProMetLeuTyrAspAlaAspGlnGlnArgIleLysPheIleLysGlnAsnMet  412

Db  1571  TCTGTGATTCCACCTATGATGTTTGATGCAGAACAAACGCAGTTCATTAAACATG  1630
                :::|||||
Qy   413  AsnGlyLeuMetAlaAlaAspProMetLysValTyrLysTyrAspArgArgGlnIleValMetAsnMetTrp  432

Db  1631  AATGGGCTTATGGAGGAGACCCTATGAAAGTGTATAAGATAGGCAGTTTATGAATGTTTGG  1690

Qy   433  SerGluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGly  452
                ::::  ||||||
Db  1691  ACTGACCATGAAAAGGAGATCTTTAAGGACAAGTTTATCCAGCATCCAAAAAACTTGGA  1750

Qy   453  LeuIleAlaSerPheLeuArgGlyLysThrValAlaGlyLysCysValLeuTyrTyrArgLeu  472
                |||||||||:::||                    :::
Db  1751  CTAATTGCATCATACTTGGAGAGGAAGAGTGTTCCTGATTGGTGTTTTTGTATTACTATTTA  1810

Qy   473  ThrLysLysAsnGlyAsnTyrLysSerLeuValAlaArgArgSerTyr---ArgArgArgGly  491
                |||||||||        |||        :::||||||
Db  1811  ACCAAGAAAAATGAGAATTATAAAGCCCTCGTCAGAGGAATTATGGGAAAGCCAGAGGGC  1870

Qy   492  LysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPro  511
                ::::::
Db  1871  AGAAACCAGCAA--------------------------  1882

Qy   512  MetProArgSerSerGlnGluGlnGluGluGlnGluAspLysGluGluLysGluGluLys  531
                :::   |||                ||:::||||||
Db  1883  ATTGCTGACCCTGCAAGAAGAAAAAGTAGAAGAAAAAAGAAGGAGAGGATAAAGCCAGAAAA  1942

Qy   532  GluGluGluLys  535
                ||:::||
Db  1943  ACAGAAAAAAA  1954
```

RESULT 7
BC035748

```
LOCUS       BC035748                1850 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens nuclear receptor co-repressor 1, mRNA (cDNA clone
            IMAGE:5749816), with apparent retained intron.
ACCESSION   BC035748
VERSION     BC035748.1  GI:23958885
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 1850)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE   22388257
   PUBMED    12477932
REFERENCE   2  (bases 1 to 1850)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: cgapbs-r@mail.nih.gov
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
```

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
  Sequencing Center (NISC),
  Gaithersburg, Maryland;
  Web site: http://www.nisc.nih.gov/
  Contact: nisc_mgc@nhgri.nih.gov
  Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
  Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
  Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
  Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
  Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
  McDowell,R., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
  Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
  Young,A., Zhang,L.-H. and Green,E.D.

  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: IRAK Plate: 79 Row: m Column: 8
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 22538460
  This clone has the following problem: retained intron.

FEATURES   Location/Qualifiers
  source   1..1850
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:5749816"
                   /tissue_type="Pancreas, Spleen, adult pooled"
                   /clone_lib="NIH_MGC_120"
                   /lab_host="DH10B"

ORIGIN   /note="Vector: pCMV-SPORT6"

Alignment Scores:
  Pred. No.:   4.6e-58   Length: 1850
  Score: 1634.00   Matches: 346
  Percent Similarity: 74.35%   Conservative: 57
  Best Local Similarity: 63.84%   Mismatches: 96
  Query Match: 12.36%   Indels: 43
  DB: 11   Gaps: 12

US-09-522-753-5 (1-2517) x BC035748 (1-1850)

Qy   16 GluProArgTyrProHisProSerLeuSerTyrProValGlnIleAlaArgThrHisThr   35
          ::: ||||||||:|||||||||||
Db   272 CAAAGTCGTTATCCTCCTCACTCTGTCCCAGTATACATTTCCCAACACCCGCCACCAGCAG 331

Qy   36 AspValGlyLeuLeuGluTyrGln------HisHisSerArgAspTyrAlaSerHisLeu   53
          :::   :||:|||||||::: |||||||||
Db   332 GAGTTCGCAGTCCCAGTTATCGTTCCTCCATCTTGAAGTGAGTCAGGCATCACAGCTT   391

Qy   54 SerProGlySerIleIleGlnProGlnArgArgProSerLeuLeuSerGluProGluGln   73
          |||||||||| ||||||||||||||||
Db   392 TTGCAGCAACAGCAGCAACAGCTTCGAAGGCGACCTTCCTGCTTTCAGAATTTCAC   451

Qy   74 ProGlyAsnGluArgSerGlnGluLeu---His-LeuArgProGluSerHis------Se   90
          ||||||||:::::|| || ||||||
Db   452 CCAGGTTCTGACAGTGTTTATTTCTTTGTATGTATTAGGCCTCAAGAAAGAGAACTAG   511

Qy   90 rTyrLeuPro---GluLeuGlyLysSerGluMetGluPhe------IleGluSerLysAr   107
          ::: ||||||:::   ::|||
Db   512 TTATGAACCGTTCATCCAGGCCCCATCCCCAGTGGATCATGATTCACTGGAATCGAAGCG 571

Qy   107 gProArgLeuGluLeuProAsp---------------------------------Pr   116
          ||||||:|| |||
Db   572 ACCACGTCTCGGAACAGGTTTCTGATTCTCATTTCAGATTTCGATTTTAACCAAGAA 631

Qy   116 oLeuLeuArgProSerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspLeuTh   136
          ||||||:|| ::: ||
Db   632 TTTAGTGCACCCGCTGCCA--------GAAGGGCTGAGGGCT--TCTGCAGATGCTAA 679

Qy   136 rLysAspArgSerLeuThrGlyLysSerLeuGlu--ProValSerProSerProProHi   155
          ||||| ::: ||| |||
Db 1757 GCAA--ATTGCTCG 1768

Db   680 GAAGGATCCAGCATTCGGAGGCAAACATGAAGCTCCATCCTCTCCAATTCGGGCGCAACC 739

Qy   155 sThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGluLeuGlnAs 175
          ::: ||||||||::: ||||||
Db   740 ATGTGGAGATGATCAAAATGCTTCACCTTCAAAACTCTCAAAGGAAGAGTTAATACAGAG 799

Qy   175 nMetAspArgValAspArgGlyIleThrMetValGluGlnIleSerLysLeuLysLyY 195
          :||||||||||||||||||::: ||||||||:::
Db   800 TATGGATCGTGTAGATCGAGAAAATGCAAAAGTAGAACAGCAGATCCTTAAACTGAAAA 859

Qy   195 sLysGlnGlnGlyLeuGluGluAlaAlaAlaLysProGluProGluLysProValSe 215
          ||||||||||||::: ||||||||||
Db   860 GAAACAACAACAGCTTGAAGAAGAGGCAGCTAAACCTCCTGAGCCTGAGAAGCCCGTGTC 919

Qy   215 rProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
          |||||||| |||||||||||||::: |||
Db   920 CCCTCCTCCTGTGGAGCAGAAACACCGCAGTATTGTCCAAATTATTTATGATGAGAATCG 979

Qy   235 gLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuPr 255
          ::: |||||||||| :::
Db   980 GAAAAAAGCAGAAGCTGCATCATAAAATTTTGAAGGTCTTGGCCCAAAGTTGAACTGCC 1039

Qy   255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAl 275
          ||||||::: |||||||||::: |||
Db 1040 ACTGTATAACCAGCCATCAGATACCCAAGTGTACCATGAGAACATCAAGACAAACCAGGT 1099

Qy   275 aMetArgLysLysLeuLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLy 295
          :::|||||||||||||||||| |||:::
Db 1100 GATGAGGAAGAAACTCATTTATTTTTAAAAGAAGAAGAACATGCAGAAAAACAAAGGGA 1159

Qy   295 sGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluAr 315
          :::||||||||||||||||||::: |||
Db 1160 ACAAAAAATCTGCCAGCGTTATGATCAGCTCATGGAGGCATGGGAGAAAAAAGTGGACAG 1219

Qy   315 gIleGluAsnAsnProArgArgArgAlaLeuLysSerLysLysValArgGluTyrTyrGluLy 335
          |||||||||| :::   :::
Db 1220 AATAGAAAATAATCCTCGGAGAAGAGCTCTAAAGAGCAAAAAAGTGAGATACTATGAAAA 1279

Qy   335 sGlnPheProGluIleLeuTyrPheLysGlnArgGluLeuGlnArgMetGlnSerArgValGl 355
          ||||||||||::: ||||||||||
Db 1280 GCAGTTTCCAGAAATTCTGAAAACAAGAGAACAGCAAAGAAAGATTTCAG---CGAGTTGG 1336

Qy   355 yGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIl 375
          ||||||||||:|||||||||| :::
Db 1337 GCAGAGGGGAGCTGGTCTTTCAGCCACCATTGCTAGGAGTGAGCATGAGATTTCTGAAAT 1396

Qy   375 eIleAspGlyLeuLeuGluGlnAsnLeuGluLysGlnMetArgGlnLeuAlaValIl 395
          ||||||||||::: |||
Db 1397 TATTGATGGGCTCTTCTGAGCAGGAGAATAATGAGAAACAAATGCGGCAGCTCTGTGAT 1456

Qy   395 eProProMetLeuTyrArgAspAlaAspProLeuAsnMetArgIleLysPheIleAsnMetGlyLe 415
          ::|||||||||:||||||::: |||
Db 1457 TCCACCTATGATGTTTGATGCAGAACAAAGAACCAGTCAAGTTCATTAACATGAATGGGCT 1516

Qy   415 uMetAlaAspProMetLysValTyrLysAspArgGlnValMetGluAsnMetTrpSerGluGl 435
          ||||||||||:::   :::
Db 1517 TATGGCAGACCCTATGAAAGTGTATAAAGATAGGCAGTTTATGAATGTTTGGACTGACCA 1576

Qy   435 nGluLysGluThrPheArgGluLysProLysAsnPheGlnHisProLysLeuLeuAl 455
          ||||||||||::: ||||||
Db 1577 TGAAAAGGAGATCTTTAAGGACAAGTTTATCCAGCATCCAAAAAAACTTTGGACTAATTGC 1636

Qy   455 aSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLy 475
          ::|||:::|||||||||||| |||
Db 1637 ATCATACTTGGAGAGGAAGAGTGTTCCTGATTGTGTTTTGTATTACTATTTAACCAAGAA 1696

Qy   475 sAsnGluAsnTyrTyrLeuValLeuValArgArgSerTyr---ArgArgGlyLysSerGl 494
          ::|||::: ||| ||||||||||
Db 1697 AAATGAGAATTATAAAGCCCTCGTCAGAAGAATTATGGGAAGCAGAGGAGGAGAAACCA 1756

Qy   494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProAr 514
          :::
Db 1757 GCAA--ATTGCTCG 1768

```
Qy    514 gSerSerGlnGluGluAspGluLysGluGluGluAlaGluLysGluGlu 534
              ||||||||     ||||||||      :::::|||||||   ||||||
Db   1769 ACCCTCGCAAGAGAGAAGTAGAACAAGAAGAAGGATAAAGCAGAAAAAACAGAAAA 1828

Qy    534 uLys 535
              :|||
Db   1829 AAAA 1832
```

RESULT 8
BC029627
LOCUS   BC029627   1835 bp   mRNA   linear   HTC 07-AUG-2002
DEFINITION Mus musculus, Similar to nuclear receptor co-repressor 1, clone
            IMAGE:4501757, mRNA.
ACCESSION   BC029627
VERSION   BC029627.1 GI:20988194
KEYWORDS   HTC.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1835)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 31 Row: k Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis. Similarity but not identity to protein
            This clone has the following problem: frame shifted.
FEATURES   Location/Qualifiers
     source   1..1835
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4501757"
                     /tissue_type="Eye, retina, mouse strain C57Bl\6"
                     /clone_lib="NIH_MGC_94"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
ORIGIN

Alignment Scores: Length: 1835
Pred. No.:   9.31e-58   Matches: 342
Score: 1626.50   Conservative: 56
Percent Similarity: 73.43%   Mismatches: 93
Best Local Similarity: 63.10%   Indels: 51
Query Match: 12.31%   Gaps: 12
DB: 11

US-09-522-753-5 (1-2517) x BC029627 (1-1835)
```

```
Qy     16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
```

```
Qy        :::
Db    282 CAAAGTCGTTATCCTTCACATTCTCGTCCAGTATACCTTTCCCAGTACCCGACACCAGCAG 341

Qy     36 AspValGlyLeuLeuGluTyrGln-----HisHisSerArgAspTyrAlaSerHisLeu 53
              ::: ::::::|| |||||||
Db    342 GAATTTGCAGTTCGACTACCGGTCTTCTCATATTGAAGTTAGCCAGGCATCACAGTT 401

Qy     54 SerProGlySerIleIleGlnProGln---ArgArgProSerLeuLeuSerGluPhe 72
              |||||||||||||||||||||||
Db    402 TTACAGCAGCAGCAGCAGCAGCAGCTTCGAAGACGACCTTCCTTGCTTTCAGAATTT 461

Qy     73 GlnProGlyAsnGlnArgSerGlnGluGluHisLeuArgProGluSer---HisSerTyr 91
              ||||||     ||||||         :::|||||
Db    462 CACCCGGGTTCTGACAGGCCTCAAGAAAGGAGAACTGGATATGAACAGTTTCACTCA--- 518

Qy     92 LeuProGluLeuGlyLysSerGluMetGluPhe------IleGluSerLysArgProArg 109
              ||| ||| :::::    ::|||||||||
Db    519 --------GGACCCTCACCGGTGGATCATGATTCCTTGGAGTCCAAGCGCCCACGC 566

Qy    110 LeuGluLeuLeuProAsp----------------------ProLeuLeu 118
              |||||||||||||                                :|::||
Db    567 CTGGAGCCAGTTTCTGATGCCCATTTCCAGCGTGTTAGTCGTGCGGTTTTACCTTAGTT 626

Qy    119 ArgProSerProLeuLeuAlaThrGlyLysGlnProAlaGly----------SerGluAspLeu 135
              |||                            |||
Db    627 CAC--------------TCGCTGCCAGAAGCTTGAGGTGTCGTCTGCAGATGCT 665

Qy    136 ThrLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProProSerProPro 154
              ||| :::                         ||||||||||||
Db    666 AAGAAGGATTCAGCATTTGGAAGCAAACATGGAAGCTCCATCCTCCTTTGGCTGGGCAA 725

Qy    155 HisThrAspProGluGluGluValProProArgLeuSerLysGluGluLeuGlu 174
              :::                            :::||||||||||
Db    726 CCATGTGGAGATGACCAAAATGCTTCACCTTCAAAGGAGGAGTTAATACAG 785

Qy    175 AsnMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLeuLys 194
              ||||||||||||||||||||
Db    786 AGTATGGACCGGGTAGACCGGAGAATTGCAAAAGTAGAGCAGCAGATCCTTAAATTGAAA 845

Qy    195 LysLysGlnGlnGlnLeuGluGlyGluGluGluAlaLeuAlaLeuLysProProGlyLysProVal 214
              |||
Db    846 AAGAAACAACAGCTAGAAGAGGAAGAGCAGCTAAGCCCCTGAACCAGAGAAGCCTGTG 905

Qy    215 SerProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn 234
              :::                           ::: |||||||||
Db    906 TCCCCTCCTCCTGGAACAAACACCGTAGTTATTGTCCAAATTATTTATTATGATGAGAAT 965

Qy    235 ArgLysAlaGluAlaAlaHisArgIleLeuGlyGlyLeuGlnValGluLeu 254
              |||||||:::||             ||||||
Db    966 CGGAAAAAGCAGAGAAGCTCATAAATAATTTGAAGGTTTGGCCCAAAGTTGAACTG 1025

Qy    255 ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGln 274
              ||||||      ||||||
Db   1026 CCACTCTATACCAGCCATCAGATACCAAGGTGTACCATGAGAACATCAAGAACAAACCAG 1085

Qy    275 AlaMetArgLysLeuLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrp 294
              |:::|||||    :::||              |||||||
Db   1086 GTGATGAGGAAAAAACTCATTTTATTTTTTAAAAGAAGAAATCATGCAAGATGAAGACAAAGG 1145

Qy    295 LysGlnLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysValGlu 314
              |||||||      ||||||||
Db   1146 GAACAAAAAATCTGCCAGCCTATGATCAGCTCATGGAAGCCTGGGAGAAAAAGTGGAC 1205

Qy    315 ArgIleGluAsnAsnProArgArgArgAlaLysSerLeuValArgGluTyrTyrGlu 334
              ||||||                         |||
Db   1206 AGAATAGAAAATAAATCCCGGAGAAGGAAAGCAAAAGAAAGGGAATACTATGAG 1265

Qy    335 LysGlnPheProGluIleArgLysGlnArgGluGluLeuGlnArgMetGlnSerArgVal 354
              |||||||                        |||||||
Db   1266 AAGCAGTTTCCAGAAATTCGAAAACAAAGAGAACAACAGAAAAGATTTCAG---CGAGTT 1322

Qy    355 GlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGlu 374
              |||||||:::                      ::|||||||||
```

```
Qy     16 GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr 35
```

```
Db    1323  GGTCAGAGGGGAGCTGGTCTTTCAGCCACCATTGCTAGGAGTGAGCATGAGATTTCTGAA  1382
                  ||||||||||||||||||||||||||||||||   |||||||   ||||||::::::
Qy     375  IleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaVal  394

Db    1383  ATTATTGATGGTCTTTCTGAACACGGAGAATAATGAGAAGCAAATGCGTCAGCTTTCTGTG  1442
                  |||||||||||||||||||||||||||||||||||||||||| |||||||:::::
Qy     395  IleProProMetLeuTyrAspAlaAspGlnArgIleLysPheLeuAsnMetAsnGly  414

Db    1443  ATTCCACCTATGATGTTGATGCAGACAAAGAAGGGTCAAATTCATCAATATGAATGGG  1502
                  |::||||||||||||||:: ::|||||||:::::||||||:::||||||:::
Qy     415  LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGlu  434

Db    1503  CTGATGGAGGATCCAATGAAGGTTTATAAAGACCAGACAGTTTATGAATGTTTGGACTGAC  1562
                  |||||||||||::||||||||||||||||||||||| |||||||||||||||||:||
Qy     435  GlnGluLysGluThrPheArgGluLysPheGlnMetHisProLysAsnPheGlyLeuIle  454

Db    1563  CATGAAAAGGAGATCTTTAAGGACAAGTTTATCCAGCATCCCAAAAAACTTTGGACTAATT  1622
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     455  AlaSerPheLeuGluLysLysThrValAlaGluCysValCysValLeuTyrTyrLeuThrLys  474

Db    1623  GCATCCTATTTGGAAAGGAAGGAGTGTTCCTGATGTGTTTATATTACTATTTAACCAAG  1682
                  |:||||||||||||:||||||   ::|||||||::|||||||||:||||||||||||
Qy     475  LysAsnGluAsnTyrLysSerGlnValArgGlySerTyr---ArgArgArgGlyLysSer  493

Db    1683  AAAAATGAGAATTATAAGGCCCCTCGTGAGAAGAGGAATTATGAAAAACGCAGAGGCAGAAAT  1742
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     494  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro  513

Db    1743  CAGCAG---------------------------------------ATTGCC  1754
                  |||||                                        :::
Qy     514  ArgSerSerGlnGluGlyLysAspGluLysGluGluLysGluLysGluGluAlaGluLysGluGlu  533

Db    1755  CGTCCCTCACAAGAGAAGAAAGTAGAAGAAGGAGGATAAAGCAGAAGAAAAAAACAGAA  1814
                  ::|||||||||| |||||||||||||||||||||||||||||||||::||||||
Qy     534  GluLys  535

Db    1815  AAAAAAA  1820
                  ::: |
```

```
RESULT 9
BC062162
LOCUS        BC062162                1819 bp    mRNA    linear    HTC 14-NOV-2003
DEFINITION   Mus musculus cDNA clone IMAGE:30240594, containing frame-shift
             errors.
ACCESSION    BC062162
VERSION      BC062162.1  GI:38328289
KEYWORDS     HTC.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 1819)
  AUTHORS    Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE      Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE    22388257
```

```
   PUBMED   12477932
REFERENCE    2  (bases 1 to 1819)
  AUTHORS    Strausberg,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (13-NOV-2003) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
             Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. David Rowe and Dr. Mina
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Genome Sequence Centre,
             BC Cancer Agency, Vancouver, BC, Canada
             info@bcgsc.bc.ca
COMMENT      Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
             Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
             Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
             Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
             Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
             Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
             Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
             George Yang, Scott Zuyderduyn, Marco Marra.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 131 Row: c Column: 2
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Hexamer frequency ORF
             analysis, Similarity but not identity to protein
             This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..1819
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30240594"
                     /tissue_type="Jaw bone, mouse, day 10.5 to 11.5 (CD-1),
                     developing maxilla and mandibula tissue containing
                     undifferentiated progenitor cells for muscle, dermis,
                     epidermis, skin, membraneous bone, cartilage and teeth"
                     /clone_lib="NIH_MGC_164"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6.1"
ORIGIN

Alignment Scores:
Pred. No.:                                    Length:            1819
Score:              1.12e-57                   Matches:           341
Percent Similarity: 1624.50                   Conservative:       59
Best Local Similarity: 73.94%                 Mismatches:         91
Query Match:         63.03%                    Indels:             50
DB:                  12.29%                    Gaps:               11

US-09-522-753-5 (1-2517) x BC062162 (1-1819)

Qy     16  GluProArgTyrProProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThr  35
               ::: |||||||||||||||||||||||||||||||||||||||||||||||||
Db    282  CAAAGTCGTTATCCTTCACATTCTGTCCAGTATACCTTTCCAGTACCCGACACCAGCAG  341

Qy     36  AspValGlyLeuLeuGluTyrGln------HisHisSerArgAspTyrAlaSerHisLeu  53
               ::: ::: ::|||||                   |||
Db    342  GAATTTGCAGTTCCTGACTACCCGTCTTCTCATATTGAGTTAGCCAGGCATCACAGCTT  401

Qy     54  SerProGlySerIleIleGlnProGln--ArgArgProSerLeuLeuSerGluPhe  72
               |||||||||||||||||||||  |||||||||||||||||||||||||||||||||
Db    402  TTACAGCAGCAGCAGCAGCAGCAGCAGCAGCTTCGAAGACGCAGCTTCCTTGCTTCAGAATT  461

Qy     73  GlnProGlyAsnGluArgSerGlnGluLeu-HisLeuArgProGluSerHisSerTyrLe  92
               |||||||||||||||||||||||||||||      |||||||  -----
Db    462  CACCCGGGTTCTGCAGGGCCTCAGGAAAAGAGAAACTGGATGAACAGTTTCACTCA-----  516
```

Qy　92 uProGluLeuGlyLysSerGluMetGluPhe------IleGluSerLysArgProArgLe 110
　　　　　　　　　　||||||| :::::::　　　　　　　||||||||||||||||
Db　517 ---------GGACCCTCACCGGTGATCATGATTCCTTGGAGTCCAAGCGCCCACGCCT 566

Qy　110 uGluLeuLeuProAsp------------------------------ProLeuLeuArg 119
　　　　||||||||||||||:::　　　　　　　　　　　　　　　　　　　　　||||||
Db　567 GGAGCCAGTTTCTTGATCGCCATTCCCAGCGTGTTAGTGCTGCGGGTTTACCTTTAGTTCA 626

Qy　119 gProSerProLeuLeuAlaThrGlyGlnProAlaGly-------SerGluAspLeuTh 136
　　　　||||||||||||　　　　　　　　　　　　　　　　　　　||||||||||
Db　627 C---------------TCGCTGCCAGAAGGCTTGAGGTCGTCTGCAGATGCTAA 665

Qy　136 rLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProSerProProHi 155
　　　　||||　　　　:::　　　　　　　　　　　　:::
Db　666 GAAGGATTCAGCATTGGAAGCAAACATGAAGCTCCATCCTCCTTTGGCTGGGCAACC 725

Qy　155 sThrAspProGluLeuGluLeuGluValProProArgLeuSerLysGluGluLeuIleGlnAs 175
　　　　||||　　　　　　：：　||||||||||||||||||||||||||||||：
Db　726 ATGTGGAGATGACCAAAATGCTTCACCTTCAAAGCTTCAAAGGAGGAGTTAATACAGAG 785

Qy　175 nMetAspArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuLysLy 195
Db　786 TATGGACCCGGTAGACCGAGATTCAAAAGTAGAGCCAGCAGATCCTTAAATTGAAAAA 845

Qy　195 sLysGlnGlnLeuGluGluGluGluValAlaLysAlaLysProProGluProGlyProValSe 215
　　　　|||||||||||||||||||||||||||||||||||||||||||||
Db　846 GAAACAACAACAGCTAGAAGAAGAGCAGCTAAGCCCCCTGAACCAGAAGAAGCCTGTGTC 905

Qy　215 rProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsnAr 235
　　　　||||||||||||||:::::||||　||||||||||||||||||||
Db　906 CCCTCCTCCTGTGGAACAAAAACACCGTAGTATTGTCCAAATTATTTATGATGAGAATCG 965

Qy　235 gLysLysAlaGluAlaAlaHisArgIleLeuGlyLeuGlyLeuGlyLeuProGlnValGluLeuPr 255
　　　　||||||||||||||||||||||||||||||||||||：：：：
Db　966 GAAAAAAGCAGAAGCAAGCTCATAAAATATTTGAAGGTCTTGGCCCAAAGTTGAACTGCC 1025

Qy　255 oLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAal 275
　　　　|||||||||||||||||||：：：
Db　1026 ACTCTATAACCAGCCATCAGATACCACCAAGTGTACCATGAGAACATCAAGACAACCAGT 1085

Qy　275 aMetArgLysLeuLeuLeuTyrPheLysSerArgArgAsnHisAlaArgLysArgGlnTrpPly 295
　　　　||||||||||　||　|||||||||||||||||||||||||||
Db　1086 GATGAGAAAACTCATTTTTTTAAAGAGAAGAAAATCATGCAAGAAAACAAAGGA 1145

Qy　295 sGlnLysPheCysGlnArgTyrArgAspGlnLeuMetGluAlaLeuGluLysValGluAr 315
　　　　:::||||||||||||||||||||||||||：：：
Db　1146 ACAAAAAATCTGCCAGCGCTATGATCAGCTCATGGAAGCGTGGGAGAAAAAAGTGGACAG 1205

Qy　315 gIleGluAsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLy 335
　　　　||||||||||||||||||||||||||||
Db　1206 AATAGAAAATAATCCCGGAGGAAAGCAAAAGCAAAACAAGGGAATACTATGAGAA 1265

Qy　335 sGlnPheProGluIleArgLysGlnLeuArgGluGlnArgMetGlnSerArgValGl 355
　　　　||||||||||||||||||||||||||||||　||||
Db　1266 GCAGTTTCCAGAAATTCGAAAACAAGAGAACAGCAGAAGAGAGATTTCAG---CGAGTTGG 1322

Qy　355 yGlnArgGlySerGlyMetSerMetSerAlaAlaArgSerGluHisGluValSerGluIl 375
　　　　|||||||||||||||||||||||||||||　　::　||||||||||
Db　1323 TCAGAGGGGAGCTGGTCTTCAGCCACCATTGCTAGGAGTGAGCATGAGGATTTCTGAAAT 1382

Qy　375 eIleAspGlyProMetLeuGlnAsnLeuGluLysGlnMetArgGlnMetAlaValI 395
　　　　|||||||||||||||||||||||：：　||||||||||
Db　1383 TATTGATGGTCTTTCTGAACAGGAGAATAATGAACAAATGCGTCAGCTTTCTGTGAT 1442

Qy　395 eProProMetLeuTyrAspAlaAspIleGlnArgIleLeuPheIleAsnMetAspGlyLy 415
　　　　||||||||||||||||||||：：||||||||||||||
Db　1443 TCCACCTATGCTATGTTTGATGCAGACAAAGAAGGGTCAAATTCATCAATATGAATGGCT 1502

Qy　415 uMetAlaAspProMetLysValTyrrIleValMetAsnMetTrpSerGluGl 435
　　　　||||　||||||||||：：||||：||||||||||||：：
Db　1503 GATGGAGGATCCAATGAAGGTTTATAAAGACACAGTTTATGAAGTTTGGACTGACCA 1562

Qy　435 nGluLysGluThrPheArgArgGluLysPheGlnHisProLysAsnPheGlyLeuIleAl 455
　　　　||||||||　|||||||||||||||||：：
Db　1563 TGAAAAGGAGATCTTTAAGGACACAAGTTTATCCAGCATCCAAAAACTTTGGACTAATTGC 1622

Qy　455 aSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLeuThrLysLY 475
　　　　||||：：||||||||||||||||||||：：
Db　1623 ATCCTATTTGGAAAGGAAGACAGTGTTCCTGATTGTGTTTATATTACTATTAACCAAGAA 1682

Qy　475 sAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArgArgGlyLysSerG1 494
　　　　|||||||||||||||||||||||||：：：：
Db　1683 AAATGAGAATTATAAGGCCCTCGTGAGAAGGAATTATGGAAAACCGAGAGGCAGAAATCA 1742

Qy　494 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetProAr 514
　　　　||||||
Db　1743 GCACGCAG--------------------------------------ATTGCCCG 1757

Qy　514 gSerSerGlnGluGlnLysGlyAspGlnLysAspGluAlaGluLysGluGlu1 534
　　　　||||||||||||||：：||
Db　1758 TCCCTTCACAAGAAGAAAAAGTAGAGAAGAGGATAAAGCAGAAAAAAAAAAAA 1817

Qy　534 u 534
Db　1818 A 1818

RESULT 10
BM909096
LOCUS　　　　　BM909096　　　　　　　　　1093 bp　　　mRNA　　　linear　　EST 12-MAR-2002
DEFINITION　AGENCOURT_6611948 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431719
　　　　　　　　5', mRNA sequence.
ACCESSION　 BM909096
VERSION　　 BM909096.1　　GI:19359475
KEYWORDS　　EST.
SOURCE　　　 Homo sapiens (human)
ORGANISM　　Homo sapiens
　　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　 1　(bases 1 to 1093)
AUTHORS　　 NIH-MGC http://mgc.nci.nih.gov/.
TITLE　　　　National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL　　 Unpublished (1999)
COMMENT　　 Contact: Robert Strausberg, Ph.D.
　　　　　　　　Email: cgapbs-r@mail.nih.gov
　　　　　　　　Tissue Procurement: ATCC
　　　　　　　　cDNA Library Preparation: Rubin Laboratory
　　　　　　　　cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
　　　　　　　　DNA Sequencing by: Agencourt Bioscience Corporation
　　　　　　　　Clone distribution: MGC clone distribution information can be
　　　　　　　　found through the I.M.A.G.E. Consortium/LLNL at:
　　　　　　　　http://image.llnl.gov
　　　　　　　　Plate: LLCM1904　row: g　column: 16
　　　　　　　　High quality sequence stop: 741.
FEATURES　　　　　　　　　Location/Qualifiers
　　　source　　　　　　　1..1093
　　　　　　　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　　　　　　　/mol_type="mRNA"
　　　　　　　　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　　　　　　　　/clone="IMAGE:5431719"
　　　　　　　　　　　　　　　/tissue_type="astrocytoma grade IV, cell line"
　　　　　　　　　　　　　　　/lab_host="DH10B (phage-resistant)"
　　　　　　　　　　　　　　　/clone_lib="NIH_MGC 99"
　　　　　　　　　　　　　　　/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
　　　　　　　　　　　　　　　EcoRI; cDNA made by oligo-dT priming. Directionally
　　　　　　　　　　　　　　　cloned into EcoRI/XhoI sites using the following 5'
　　　　　　　　　　　　　　　adaptor: GGCACGAG(G). Library constructed by Ling Hong
　　　　　　　　　　　　　　　in the laboratory of Gerald M. Rubin (University of
　　　　　　　　　　　　　　　California, Berkeley) using ZAP-cDNA synthesis kit
　　　　　　　　　　　　　　　(Stratagene) and Superscript II RT (Life Technologies).
　　　　　　　　　　　　　　　Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:　　　6.68e-58　　　　　　Length:　　　1093
Score:　　　　　1624.00　　　　　　Matches:　　　331

```
Percent Similarity:      92.78%      Conservative:      3
Best Local Similarity:   91.94%      Mismatches:       14
Query Match:             12.29%      Indels:           12
DB:                      12          Gaps:              3
```

US-09-522-753-5 (1-2517) x BM909096 (1-1093)

```
Qy   2151 HisProGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer 2170
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     19 CACCCACAGCAGCTCAGCGACCACCCTGCCCCCGCCCCCCTACTCCTTCCCTGGGCCAGC 78

Qy   2171 CysProValLeuAspLeuArgArgProSerLeuTyrLeuProProProAspHis 2190
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79 TGCCCCGTCCTGGACCTCCGCCGCCCACCAGTGACCTTCTACCTCCCGCCCCCGGACCAT 138

Qy   2191 GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsn 2210
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    139 GGTGCCCCGGCCCGTGGCTCCCCCACAGCGAAGGGGCAAGAGGTCTCCAGAGCCAAAC 198

Qy   2211 LysThrSerValLeuGlyGlyGlyGluAspGlyProValSerProProGluGlyGly 2230
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 AAGACGTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTCCCCACCGGAGGGC 258

Qy   2231 MetThrGluProGlyHisHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlu 2250
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    259 ATGACGGAGCCAGGGCACTCCCGGAGTCGTGTGTACCCGCTGCTGTACCGGGATGGGGAA 318

Qy   2251 GlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla 2270
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    319 CAGACGGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCC 378

Qy   2271 PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIle 2290
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 TTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATC 438

Qy   2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnInProGly 2310
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 AACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACATATCCCAGCCAGCCTGGG 498

Qy   2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 ACGGAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCAG 558

Qy   2331 AlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgGlyAlaLeu 2350
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    559 GCGGTGCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTC 618

Qy   2351 MetGlyLysTyrAspGlnInTrpGluSerProProLeuSerProProLeuSerAlaPheAsnAlaPro 2370
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 ATGGGTAAATATGACCCAGTGGAGGAGTCCCCGCTCAGCGGAGTTTTTTAACCCT 678

Qy   2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgGlySer 2390
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    679 CTGAATGCCAGTGCCAGCCTGCCGGCTGCTATGCCCATAACCGCTGCTGCGACGGGAGT 738

Qy   2391 AspHisThrLeuThrSerProGlyLysGlyLyGlyLysAlaLysValSerGlyGlyArgProSer 2410
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 GACCACCACACTCACCTCGCCAGGTGGCGGGGAAAGCCAAGGTCTCTGGCAGACCCAGC 798

Qy   2411 SerArgGlySerAlaLysSerAlaProGlyLeuAlaSerGlyValAspArgProProSerVal 2430
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    799 AGCCGAAAAGCCAAGTCCCGGCNCCCGGCCATCTGGGACCCGGCCCACCCTCTGTC 858

Qy   2431 SerSerValHis--SerGluGlyGlyAspCys-AsnArgArgThrProLeu--ThrAsnArgVal 2449
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    859 TCCTCAGTGCACTTCGGAGGGAGACTGGAAAACCCGCGCTCAACAAACCGGCTG 918

Qy   2450 -TrpGlu--AspArgProSerSerAlaGly-SerThrPro--PheProTyrAsnProLeuIl 2468
          :::||||||||||||||||||||||||||||||||||||::::
Db    919 GTGGGAACGACACAGGCCCTTTCTCCGGCAGGTTTCCCCTACACCCCCTGGAT 978

Qy   2468 eMetArgArgLeuGlnInAlaGlyValMet------AlaSerProProGlyLeuProAl 2486
          ||||||::::|||        :::
Db    919 CAAAGTCGTTATCCTTCACATTCTGTCCAGTATACCTTTCCAGTACCGACACCACAGCAG 549
```

```
Db   1591  ATTATTGATGGTCTTTCTGAACAGAGAGAATAATGAGAAGCAAATGCGTCAGCTTTCTGTG 1650
                ||||||||||||||||      |||||||||||||   ||:|||  |||||||||||||||
Qy    395  IleProProMetLeuTyrAspAlaAspLProAsnArgIleLysPheIleAsnMetAsnGly 414
Db   1651  ATTCCACCTATGATGTTTGATGCAGAACAAAGAGGGTCAAATTCATCAATATGATAGGG   1710
                ||||||||||||||||||||||   |:::|||||      ||:|||  ||||||||||||
Qy    415  LeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGlu  434
Db   1711  CTGATGGAGGATCCAATGAAGTTTATAAAGACAGAGTTTATGAATGTTTGGACTCAC    1770
                |||||||||||||||||||||||||||||||||:::|||||  |||||||||||||:::
Qy    435  GlnIleLeuGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIle  454
Db   1771  CATGAAAAGGAGATCTTTAAGGACAAGTTTATCCAGCATCCAAAAAACTTTGGACTAATT 1830
                ||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||:
Qy    455  AlaSerPheLeuGluArgLysThrValAlaGlucValGluLeuTyrTyrLeuTyrThrLys 474
Db   1831  GCATCCTATTTGGAAAGGAGAGAGTTCCTGAGTGGTTTTTATATTACTATTTAACCAAG  1890
                |||||||||||||||||||||||||||:|||||||||  ||||||||||||||||||||
Qy    475  LysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArgArgGlyLysSer 493
Db   1891  AAAAATGAGAATTATAAGGCCCTCGTGAGAGAGGAATTATGGAAAACGCAGAGGCAGAAAT 1950
                |||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Qy    494  GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProMetPro   513
                                                                   :::
Db   1951  CAGCAGCAG-------                                        ---ATTGCC 1965
Qy    514  ArgSerGlnGluGlyLysAspGluLys 523
                |||||||   :::|||
Db   1966  CGTCCCTCACAAGAAGAAGAAAAAAAAAAAAAA 1995
```

```
RESULT 12
BM555371
LOCUS       BM555371          965 bp      mRNA    linear    EST 20-FEB-2002
DEFINITION  AGENCOURT 6577786 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470428
            5', mRNA sequence.
ACCESSION   BM555371
VERSION     BM555371.1  GI:18795772
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 965)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1977 row: d  column: 13
            High quality sequence stop: 670.
FEATURES             Location/Qualifiers
     source          1..965
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5470428"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_41"
                     /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCAGGA(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
```

```
Db    550  GAATTTGCAGTTCCTGACTACCGCGTCCTTCTCATATTGAAGTTAGCAGGCATCACAGCTT 609
                :::        :::|||       |||
Qy     54  SerProGlySerIleIleGlnProGln---ArgArgArgProSerLeuLeuSerGluPhe  72
Db    610  TTACAGCAGCAGCAGCAGCAGCTTCGAACGACGTTCCTTCCTTGCTTTCAGAATTT      669
                ||||||||||||:::||||
Qy     73  GlnProGlyAsnGluArgSerGlnLeuHisLeuArgProGluSer---HisSerTyr     91
Db    670  CACCCGGGTTCTGACAGGCCCTGACAAGAGGAAACTGGAATGAACAGTTCACTCA---   726
                |||||||:::|||
Qy     92  LeuProGluLeuGlyLysSerGluMetGluPhe------IleGluSerLysArgProArg 109
Db    727  --------GGACCCTCACCGGTGATCATGATTCCTTGGAGTCCAAGCGCCCACGC       774
                            :::|||                         ProLeuLeu   118
Qy    110  LeuGluLeuProAsp-------                                      ProLeuLeu 118
Db    775  CTGGAGCCAGTTTCTGATGCCATTCCAGCGTGTTAGTGTCGTGCCGGTTTTACCTTTAGTT 834
                |||||||||||||||
Qy    119  ArgProSerProLeuLeuAlaThrArgGlyGlnProAlaGly--------SerGluAspLeu 135
Db    835  CAC-----                                            TGCCTGCCGAGAGGCTTGAGGTCGTCTGCAGATGCT 873
                |||||
Qy    136  ThrLysAspArgSerLeuThrGlyLysLeuGlu---ProValSerProSerProProPro 154
Db    874  AAGAGGATTCAGCATTGGAAGCAAACATGAAGCTCCATCCTCTCCTTTGGCTGGGCAA   933
                ||||||    :::
Qy    155  HisThrAspProGluLeuGluLeuValProProArgLeuSerLysGluGlnLeuGluGln 174
Db    934  CCATGTGGAGATGACCAAAATGCTTCACCTTCAAAGCTTCAAAGGACGAGTTAATACAG   993
                ||||||
Qy    175  AsnMetAspArgValAspArgGluIleThrMetValGluGlnIleSerLysLeuLys     194
Db    994  AGTATGGACCCGGGTAGACCGGATTGCAAAGATTGCCTTAAATTGAAA          1053
                :::|||||||||||||
Qy    195  LysGlnGlnGlnLeuGluGlnGlnAlaAlaLysAlaProGluProGlyLysProVal    214
Db   1054  AAGAAACAACAGCTAGAAGAAGAAGCAGCAGCTAGCTGAACCAGAGAAGCCTGTG       1113
                |||||||||||||||
Qy    215  SerProProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn 234
Db   1114  TCCCCTCCTGTGAACAAAAAACCACCGTAGTATTGTCCAAATTATTTATGATGAGAAT    1173
                |||||||||||||||
Qy    235  ArgLysLysAlaGluAlaAlaHisArgIleLeuGluLeuGlyLeuProGlnValGluLeu 254
Db   1174  CGGAAAAAAGCAACGAAGAAGCCTCATAAAATATTTGAAGGTCTTGGCCCAAAAGTTGAACTG 1233
                |||||||||||||||
Qy    255  ProLeuTyrArgAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysAsnGln 274
Db   1234  CCACTCTATAACCAGCCATCAGATACCAAGGTACCATGAGAACATCAAGACAAACCAG     1293
                |||||||||||||||
Qy    275  AlaMetArgLysLysArgLeuTyrPheLysArgArgAsnHisAlaArgLysSerGlnTrp 294
Db   1294  GTGATGAGGAAAAAAACTCATTTTATTTTTAAAAGAGAATCATGCAAGAAAACAAAAGG  1353
                |||||||||||||||
Qy    295  LysGlnLysPheCysGlnTyrArgAspGlnLeuMetGluAlaLeuGluLysValGlyGlu 314
Db   1354  GAACAAAAAATCTGCCAGCGCTATGATCAGCTCATGACCGAAGCGTGGGAGAAAAAGTGGAC 1413
                |||||||||||||||
Qy    315  ArgIleGluAlaAsnAsnProArgArgArgArgAlaLysGlyGluSerLysValArgGluTyrGlu 334
Db   1414  AGAATAGAAAAATAATCCCCGGAGAAAGCAAAAGAAACAGGGAATACTATGAG        1473
                |||||||||||||||
Qy    335  LysGlnIlePheProGluIleArgLeuLeuGlnArgGluMetGlnSerArgArgVal    354
Db   1474  AAGCAGTTTCCAGAAATTCGAAATCGAAACAAAGACAGCAGAAAGAATTTCAG---CGAGTT 1530
                |||||||||||||||
Qy    355  GlyGlnArgGlyLeuGlyLeuSerMetSerSerAlaAlaArgSerGluHisGluValSerGlu 374
Db   1531  GGTCAGAGGGGAGCTGGTCTTTCAGCCACCATTGCTAGGAGTGAGCATGAGATTTCTGAA   1590
                |||||||||||||||
Qy    375  IleIleAspGlyTyrLeuSerGluGlnGlnLeuAsnLeuGluLysGlnIleMetArgIleLeuAlaVal 394
```

```
Qy    2480  oProProGly---LeuProAlaGly------SerGlyProLeuAlaGlyProHisHis 2496
            ||||||||||        ||||||||       |||   ||||||
Db     903  CCCACCGGGGCCTTCCCCGCGGGGGGCAGCAGGCGGGGGGGGCCCCCTCGGCTGGGGCCCCCAC 960

RESULT 13
LOCUS       BM553310                1126 bp    mRNA    linear    EST 20-FEB-2002
DEFINITION  AGENCOURT_6572642 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5467163
            5', mRNA sequence.
ACCESSION   BM553310
VERSION     BM553310.1  GI:18791930
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1126)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1968  row: 1   column: 12
            High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1..1126
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5467163"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_41"
                     /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:                                Length:              1126
Score:              2.08e-54              Matches:             314
Percent Similarity: 1540.00               Conservative:        9
Best Local Similarity: 82.19%             Mismatches:          16
Query Match:        79.90%                Indels:              55
DB:                 11.65%                Gaps:                1
                    12

US-09-522-753-5 (1-2517) x BM553310 (1-1126)

Qy    1457  LeuLysTyrAspThrGlyAlaSerThrThrGlySerLysHisAspValArgSerLeu 1476
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2  CTCAAGTACGACACCGGCGCCGTCCACCACTGGCTCCAAAAGCACGACGTACGGTCCCTC 61

Qy    1477  IleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAla 1496
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62  ATCGGCAGCCCCGGCCGGACGGTCCCACCCCTGCACCCCGTGGATGTGATGGCCGACGCC 121

Qy    1497  ArgAlaLeuGluArgAlaCysTyrGlyGluLeuLysSerArgProGlyThrArgAlaSer 1516
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122  CGGGCACTGGAACTGCCTGCTACGAGGAGAGCCGGACCGGACCGCCAGC 181

Qy    1517  SerSerGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLeuGlyLysPro 1536
```

```
ORIGIN        NIH_MGC Library."

Alignment Scores:
Pred. No.:                                Length:              965
Score:              2.86e-55              Matches:             304
Percent Similarity: 1559.00               Conservative:        0
Best Local Similarity: 95.00%             Mismatches:          11
Query Match:        95.00%                Indels:              5
DB:                 11.80%                Gaps:                3
                    12

US-09-522-753-5 (1-2517) x BM555371 (1-965)

Qy    2182  AspLeuTyrLeuProProAspHisGlyAlaArgGlySerProHisSerGlu 2201
            ||||||||||||||||||||||||||||||||||||||||||||||
Db       3  GACCTCTACCTCCCGCCCCCCGGACCATGGTGCCCCGGCCCCCCACAGCGAA 62

Qy    2202  GlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGly 2221
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63  GGGGGCAAGAGGTCTCCAGACCAAACAAGACGTCGGTCTTGGGTGGTGAGGACGGT 122

Qy    2222  IleGluProValSerProProGluGlyMetThrGlyHisSerArgSerAlaVal 2241
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     123  ATTGAACCTGTGTCCCACCGGAGGGCCATGACCGGACCAGGGCACTCCCGGAGTGCTGTG 182

Qy    2242  TyrProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSer 2261
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183  TACCCGCTGCTGTACCGGGATGGGGAACAGACGGAGCCCAGCAGGATGGGCTCCAAGTCT 242

Qy    2262  ProGlyAsnThrSerGlnProAlaPheProSerLysSerThrGluSerAsnSerAla 2281
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db     243  CCAGGCAACACCAGCCAGCCCGCCTTCTTCAGCAAGAGCACTGAGAGCAACTCCGCC 302

Qy    2282  MetValLysSerLysSerGlnGluIleLeuAsnLysLysLeuAsnThrHisAsnArgAsnGlu 2301
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303  ATGGTCAAGTCCAAGAGCCAAGAGATCAACAAGAAGCTGAACACCCACAACCGGAATGAG 362

Qy    2302  ProGluThrArgAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     363  CCTGAATACAAATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACCGGA 422

Qy    2322  ThrGlyLeuMetThrTyrTyrArgSerGlnAlaValGlnHisAlaSerThrAsnMetGly 2341
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     423  ACAGGCCTTATGACCTATAGAAGCCAGGCGGTCAGAACATGCCAGCACCAACATGGGG 482

Qy    2342  LeuGluAlaIleIleLeuArgGlyAlaLeuMetGlyLysTyrLysTyrAspGlnTrpGluSerPro 2361
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483  CTGGAGGCCATAATTAGAAAGGCCACTCATGGGTAAAATATGACCAGTGGGAAGAGTCCCCG 542

Qy    2362  ProLeuSerAlaAlaAsnAlaPheAsnAlaSerAlaSerAlaSerLeuProAlaAlaMet 2381
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     543  CCGCTCAGCGCCAATGCTTTTAACCCTCTGAATGCCAGTGCCAGCTGCCGCTGCTATG 602

Qy    2382  ProIleThrAlaAlaAspArgArgSerAspHisThrLeuThrSerProGlyGlyGlyGly 2401
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db     603  CCCATAACCGCTGCTGACGGACCGAGTGACACCACACTCACCTCGCCAGGTGGCGGCGGG 662

Qy    2402  LysAlaLysValSerGlyArgProSerSerArgLysLysSerProAlaProGlyLeu 2421
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     663  AAGGCCAAGGTCTCTGGCAGACCCAGCAGCAGGAAAAGCCAAGTCCCCGGCCCTG 722

Qy    2422  AlaSerGlyAspGlyArgProSerValSerSerValHisSerGlyGluGlyGlyGlyGly 2441
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     723  GCATCTGGGGACGGGGACCGGCCACCCGTCTCCTCAGTGCACTCGGAGGGAGACTCAACCGA 782

Qy    2442  ArgThrProLeuThrAsnArgValTrpGluArgProSerSerArgGlySerThrPro 2461
            |||||||||||||||||||||||||||||||||||||||||||||
Db     783  CGGACGCCGCTCACCAACCGGGTGTGGGAGGACAGGCCCTCGTCCGGGACAGCCA 842

Qy    2462  PheProTyrAsnProLeuIleMet--ArgLeuGlnAlaGly---ValMetAlaSerProPr 2480
            ||||||||||||||||       ||||||||
Db     843  TTCCCCTACAAACCCCTGATCATGCNNGCTGCANNGGCGGNTGGTCATGGTTTCCCACC 902
```

```
VERSION      BM477568.1  GI:18526610
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1057)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12272  row: p  column: 05
             High quality sequence stop: 674.
FEATURES             Location/Qualifiers
     source          1..1057
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5554420"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_85"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:                           Length:              1057
Score:             5.63e-52          Matches:             308
Percent Similarity: 1480.50          Conservative:        2
Best Local Similarity: 87.57%        Mismatches:          27
Query Match:       87.01%            Indels:              20
Db:                11.20%            Gaps:                6
                   12

US-09-522-753-5 (1-2517) x BM477568 (1-1057)

Qy    929 AlaGluGlyGlyAlaAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThr 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCGAGGGCGGCGCCGACAAGAACCGGCTCCTGTCCCCAAGGCCCAGCCCTTCCTCACCCCGACT 60

Qy    949 GlyAspProArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysProLeuLysLeuLysGln 968
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGGCGACCCCCGGGCCAATGCCTCACCCCAGAAGCCACTGGACCTGAAGCAGCTGAAGCTGAAGCAG 120

Qy    969 ArgGlaAlaAlaAlaIleProProIleGlnValThrLysProProProProProProGlyGluAsp 988
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGAGCGGCTGCCATCCCCCCATCCAGGTCACCAAAGTCCATGAGCCCCCCCCCCGGGGAGGAC 180

Qy    989 AlaAlaProThrLysProAlaProProAlaProProProGlnAsnLeuGlnProGlu 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCAGCTCCCACCAAGCCAGCTCCCCCAGCCCCACCGCCCCAAAACCTGCAGCCGGAG 240

Qy   1009 SerAspAlaProGlnProGlyGlySerProArgGlyAlaGlyLysSerArgSerProAlaPro 1028
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGCGACGCCCCCCAGCCAGGCGGCAGCCCCCGGGGCAAGAGCAGGAGCCCCGGCACCC 300

Qy   1029 ProAlaAspLysGlu-----------------------------------AlaPheAlaAlaAlaGluAlaGln 1040
          |||||||||||||                                    ||||||||||||||||||||||
Db    301 CCCGGCGACAAGGAGGAGAGAAGCCTGTGTCTTCGCAGCCTTCGCAGCCTCCGCAGCCCGGACCTGGAGGCCCAGCAG 360

Qy   1041 LysLeuProGlyLysProAspProProCysTrpThrThrSerGlyLeuProPheProValProProArg 1060
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```
```
Db    182 AGCTCGGGGGGCTCCATTGCCGCGGCGCGCCCGGTCATTGTGCCTGGCCAGCCG 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1537 ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisProLeuProArg 1556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 AGCCAGAGCCCCTAACCTATGAGGACCACGGGGCACCCCTTTGCGCGCCACTCCCACGA 301

Qy   1557 GlySerProValThrMetArgGluProThrProArgLeuGlnLeuGluGlySerLeuSerSer 1576
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 GGTTCGCCCGTGACCACGCGGGAGACCCGCGCACGCGAGGAGCAGCTTTCGTCC 361

Qy   1577 SerLysAlaSerGlnAspArgGlyLysLeuThrSerThrProArgGluIleAlaLysSerPro 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 AGCAAGGCATCCCAGGACCGAAAGTCGACGTCGCAGGCCTCGTGAGATCGCCAAGTCCCCG 421

Qy   1597 HisSerThrValProGluHisHisProHisSerProIleSerProTyrGluHisLeuLeuArg 1616
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 CACAGCACCGTGCCCGAGCACCACCCACACCCCATCTCGCCCTATGAGCACCTGCTTCGG 481

Qy   1617 GlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrThrSer 1636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCCCTGGCCTTCGACCCCACCTCC 541

Qy   1637 IleProProArgGlyIleProLeuGluAspProAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAla 1656
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 ATACCCCCGGCGCATCCCCTCTGGACGCAGCCGCGTGCCTACTACCTGCCCCGACACCTGGCC 601

Qy   1657 ProAsnProThrTyrProHisTyrProTyrLeuIleArgGlyTyrArgTyrProAspThr 1676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 CCCAACCCCACCTACCCGCACTACCCCCACCCTTCATCCGCGGTACCCGCACACG 661

Qy   1677 AlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnMetHis 1696
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 GCGGCGCTGGAGAACCGGCAGACCATCATCAATGACTACATCACCTCGCAGATGCAC 721

Qy   1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArg 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CACAACACGGCCACCGCCATGGCCCAGCGAGCTGATATGCTGAGGGGGCCTCGCCCCGC 781

Qy   1717 GluSerSerLeuAlaLeuAsnTyrAlaAlaGlyTyrProArgGlyIleIleAspLeuSerGln 1736
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    782 GAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCCGA-------------- 820.

Qy   1737 ValProHisLeuProValLeuValProGlyThrProGlyThrProAlaThrAlaMetAsp 1756
          --------------------------------------------------
Db    820 -------------------------------------------------- 820

Qy   1757 ArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerPro 1776
          --------------------------------------------------
Db    820 -------------------------------------------------- 820

Qy   1777 LeuSerProGlyGlyProThrHisThrLysProThrThrThrSerSerGluArg 1796
          ||||||||    ||||||||||
Db    821 ------------GGAGGTCCGACACCCTTGACAAAACCACC--CCCACGTCCTCTGTCCGAACGG 870

Qy   1797 GluArgAspArg-AspArgGluArgAsp-ArgGluArgGluGluLysSerIleLeuLeuT 1816
          |||||||||   ::|::::||:  :||||||||||
Db    871 GAGCCAGACCGGGGATCCAAAGCGGGAACCCGAATTCGCACGGCGGGAAAAGTCCATCCTCC 930

Qy   1816 hrSerThrThrVal-GluHisGlyAlaProIleTrp-ArgProGlyThrGluGlnSerGlnSerSe 1835
          ||||| |||  :||||||| :|  :||||||||
Db    931 CGTTCCACCACGACGTGGGAGCCCGACCCCTGGAAAACCTGTAACGGAGCCAAACAA 990

Qy   1835 rGlySerGlySerGlyGlyGlyGlyGly 1845
          :||||||||||    :||||||
Db    991 CGGAAGCATCCCGGGAAACTGGGGGGGGGGGT 1021

RESULT 14
BM477568
LOCUS       BM477568   6484946   1057 bp    mRNA     linear    EST 05-FEB-2002
DEFINITION  AGENCOURT_6484946 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5554420
            5', mRNA sequence.
ACCESSION   BM477568
```

```
FEATURES             Location/Qualifiers
     source          1..1051
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5516745"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_41"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     SuperScript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:                          1..1051                 Length:              1051
Score:                              6.15e-52                Matches:             305
Percent Similarity:                 1479.50                 Conservative:        10
Best Local Similarity:              78.95%                  Mismatches:          24
Query Match:                        76.44%                  Indels:              60
DB:                                 11.20%                  Gaps:                4
                                    12

US-09-522-753-5 (1-2517) x BM423558 (1-1051)
```

```
Qy  1457  LeuLysTyrAspThrGlyAlaSerThr-GlySerLysLysHisAspValArgSerLe  1476
                |||||||||||||||||||||:::||    ||||||||||||||||
Db     2  CTCAAGTACGACTCCGGGGCCGTCCACCACTGGGCTCCAAAAGCACGAGTACGGTCCCT   61

Qy  1476  uIleGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAl  1496
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  CATCGGCAGCCCCGGACGTTCCCACCCGTGCACCCGCTGGATGTGATGGCCGACGC    121

Qy  1496  aArgAlaLeuGluArgAlaCysTyrGlyGluSerLeuLysSerArgProGlyThrAlaSe  1516
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   122  CCGGGCACTGGAACGTGCCTGCTACGAGGAGAGCCCGGCCAGGGACCGGCCAG   181

Qy  1516  rSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProPheAlaProHisPr  1536
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   182  CAGTCTCGGGGGGGTCCATTGCCCGGGGCGCCCCGGTCATTGTGCCTGAGCTGGGCAAGCC   241

Qy  1536  oArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProAr  1556
                |||||||||||||||||||||||||||||||||||||||||||||||
Db   242  GCGGCAGAGCCCCCTAACCTATGAGGACCACGGGCACCCTTTGCCGGGCCACCTCCACG   301

Qy  1556  gGlySerProValThrMetArgProThrProArgLeuGlnGluGlySerLeuSerSerSe  1576
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   302  AGGTTCGCCCGTGACCACGCGGGAGCCCGCCCGCTGCAGGAGGGGCAGCCTTTCGTC   361

Qy  1576  rSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerPr  1596
                |||||||||||||||||||||||||||||||||||||||||||||||
Db   362  CAGCAAGGCATCCCAGGACCCAGAAAGTCACGTCGATCGCAGCCAAGTCGCCAAGTCCCC   421

Qy  1596  oHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisGluLeuAr  1616
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   422  GCACAGCACCGTGCCCGAGCACCACCCACACCCCATCTCGCCCTATGAGCACGTGCTTCG   481

Qy  1616  gGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuLeuAlaPheAspProThrSe  1636
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   482  GGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCCCTGCCTGCCTTCGACCCCACCTC   541

Qy  1636  rIleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrLeuProArgHisLeuAl  1656
                ||||||||||||||||||||||||||||||||||||||||||||||
Db   542  CATCCCCCGGGGCATCCCTCTGGACGCGGCAGCCGCCGCTTACCTGCCCCGACACCTGGC   601

Qy  1656  aProAsnProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspTh  1676
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db   602  CCCCAACCCCACTTACCCCACCTGTACCCACCTACCCACCTACCTGATCCGCGGCTACCCCGACAC   661
```

```
Db   361  AAGCTGCCTGGGGACCCCCTTGCTGGACTTCCGGCCTGCCCCCGTGCCCCCCGT  420

Qy  1061  GluValIleLysAlaSerProHisAlaAspProSerAlaPheSerPheTyrAlaProPro  1080
                |||||||||||||||||||||||||||||||||||||||
Db   421  GAGGTGATCAAGGCCTCCCCCCGACCCCTCCAGCCCTTCTCCTTCTACGCTCCACCT  480

Qy  1081  GlyHisProLeuProLeuGlyLeuHisHisAspProValLeuProValProArgProPro  1100
                |||||||||||||||||||||||||||||||||||||||
Db   481  GGTCACCCCTGCCCCTGGCCTCCATGACACTGCCCCGCCCGTCCCGCCGCCCACCC  540

Qy  1101  ThrIleSerAsnProProLeuLeuSerSerAlaLysHisProSerValLeuGluArg  1120
                |||||||||||||||||||||||||||||||||||||||
Db   541  ACCATCTCCAACCCGCCCTCCCCCATCTCCTCTGCCAAGCACCCCAGCGTCCTCGAGAGG  600

Qy  1121  GlnIleGlyAlaIleSerGlyMetGlyValGlnLeuHisLeuValProTyrSerGluHis  1140
                |||||||||||||||||||||||||||||||||||||||
Db   601  CAAATAGGTGCCATCTCCAAGGAATGTCGGTCCACGTCCCGTACTCAGAGCAT  660

Qy  1141  AlaLysAlaProValGlyThrMetGlyLeuProLeuProMetAspProLysLeuLys  1160
                |||||||||||||||||||||||||||||||||||||||
Db   661  GCCAAGGCCCCGGTGGGCCCTGTCACCATGGGGCTGCCCCTGCCCATGGACCCCAAAAAG  720

Qy  1161  LeuAlaProPheSerGlyValValLysGlnLeuGlnLeuSerProArgGlyGlnAlaGlyPro  1180
                |||||||||||||||||||||||||||||||||||||||
Db   721  CTGGCACCCTTCAGCGGCGAGTGAAGCAGGAGCAGCTGTCTCCCACGGGGCCAGGCTGGGCCA  780

Qy  1181  ProGluSerLeuGlyValProThrArgAlaGlnGluAlaSerValLeuArgGlyThrAlaLeu  1200
                ||||||||||||||||||||||||||||||||||||||||
Db   781  CGGGAGAG-CTGGGGGGTGCCACA-GCCCAGGAGCGTCCGTGCTGAGAGGGACAGCTCTG  838

Qy  1201  GlySerValProGlyGlySerIleThrLysGlyGlyIleProSer---ThrArgValProSer  1219
                |||||     ||
Db   839  GGCTCAGNTTCCGGCGGAGCATCACCAAAGGCATTCCCAGCACAGGGGGTCCCTCGG  898

Qy  1220  AspSer-AlaIleThrTyrArg-GlySerIleThrHisGly---ThrProAlaAspValL  1238
                ||||||           |||||
Db   899  GACAGCGNCCATCACATACCGCGGGTTCCATCACCACGGNNACGCCCAGCTGGACGTCC  958

Qy  1238  euTyrTyrGlyThr---IleThrArgIle---IleGlyGlyAspSerProSerArgLeuA  1256
                ||||||           :::||
Db   959  TGGA-CAAGGCNCACCCATCACCAGGNATCATCGGGCGAAGGACAGNCCGGAGTCGCTTG  1017

Qy  1256  spArg--------GlyArgGluAspSerLeuProLys  1265
                ||||||
Db  1018  GAAACNCCCGNCCGGAAGGAACAGCCCCTGGCCCAAA  1055
```

```
RESULT 15
BM423558
LOCUS       BM423558              1051 bp    mRNA      linear    EST 29-JAN-2002
DEFINITION  AGENCOURT 6399411 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5516745
            5', mRNA sequence.
ACCESSION   BM423558
VERSION     BM423558.1  GI:18391770
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1051)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2017 row: n  column: 10
            High quality sequence stop: 667.
```

```
Qy   1676 rAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHi 1696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 GGCGGCGCTGGAGAACCGCAGACCATCATCAATGACTACATCACCTCGCAGCAGATGCA 721

Qy   1696 sHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProAr 1716
          |||||||||||||||||| ||||||||| ::|||||||||||||||||||||||||||||
Db    722 CCACAACACCGGCCACCGCCACCGCCAGCGAGCTGATGCTGAGGGGTCTCCGCCCG 781

Qy   1716 gGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyLeuProArgGlyIleIleAspLeuSerGl 1736
          ||||||||||||||||||||||||||||||||||||||||
Db    782 CGAGTCCTCGCTGGCACTCAACTACGCTGCGGGTCCC------------------------ 818

Qy   1736 nValProHisLeuProValLeuProThrProGlyThrProAlaThrAlaMetAs 1756
          ------
Db    818 ------------CCGAGGAGGCCAACCTTGACCAAACCACCCCAGTCCCTGTCCGAG 818

Qy   1756 pArgLeuAlaTyrLeuProThrAlaProGlnProPheSerArgHisSerSerSerPr 1776
Db    818 ------

Qy   1776 oLeuSerPro-GlyGlyProThrHis-LeuThrLysProThrThrThrSerSerGlu 1795
          |||||||||||                                |||||
Db    819 ------CCGAGGAGGCCAACCTTGACCAAACCACCCCAGTCCCTGTCCGAG 871

Qy   1796 ArgGluAspArg---AspArgGluArgAspArgAspArgGlu-ArgGluLysSerIl 1814
          ||||||| :::||    :::                |||||||
Db    872 CGGGAGGACGAGAACCGGGAGATCCGGGATCGGGAACCGGGAAAGGTCCAT 931

Qy   1814 eLeuThr---SerThrThrValGluHisAlaProIleTrp-ArgProGlyThrGluG 1833
          ||||                 :::||||||||||||     :::||||||||||||
Db    932 CCTTCCGTTCCCCAACGGAGGAGCACGCCCCCATCTGGAAAACCTGGTACCCAAC 991

Qy   1833 lnSerSerGlySerGly-----------SerGlyGlyGly 1845
          ::: ||||||||||                |||||||||||
Db    992 CCAAACAACGGGCACCAGGGCACTTACACATTGCGGGGGGGGGG 1038

Db    818 ------
```

```
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

ORIGIN

```
Alignment Scores:
Pred. No.:                              Length:               937
Score:                1.16e-51          Matches:              294
Percent Similarity:   1471.50           Conservative:           2
Best Local Similarity: 77.28%           Mismatches:             7
Query Match:          76.76%            Indels:                80
DB:                   11.14%            Gaps:                   3
                      13
```

US-09-522-753-5 (1-2517) x BQ949384 (1-937)

```
Qy   2127 LysGlyHisGlnArgValValThrLeu-AlaGlnHisIleSerGluValIleThrGlnAs 2146
          :::|||                    ||| |||||||||||||||||||||||||||||||
Db     20 CGAGGGTCCCAGCGGCGGTGGTCACCCTGGGCCCAGCACATCAGTGAGGTCATCACACAGGA 79

Qy   2146 pTyrThrArgHisHisProGlnLeuLeuSerAlaProLeuProLeuTyrSerPh 2166
          |||||||||||||||||||||||||||||||||||||||||::|||||||||||
Db     80 CTACACCCGGCACCACCACCCACAGCAGTCAGCCCCGCCCCCCTCTACTCCTT 139

Qy   2166 eProGlyAlaSerCysProValLeuAspLeuArgArgProSerAspLeuTyrLeuPr 2186
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    140 CCCTGGGGCCCAGCTGCCCCGTCCTGGACCTCCGCCGCCCACCCAGTGACCTCTACCTCCC 199

Qy   2186 oProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSe 2206
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    200 GCCCCCGGACCATGGTGCCCCGGCCCGTGGCTTCCCCCACAGCGAAGGGGCAAGAGGTC 259

Qy   2206 rProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluGluProValSe 2226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    260 TCCAGAGCCAAACAAGACGTCGGTCTTGGGTGGTGGTGATTGAACCTGTGTC 319

Qy   2226 rProProGluGlyMetThrGluProGlyHisSerArgSerAlaValProLeuLeuTy 2246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 CCCACCGGAGGGCATGACGGAGCCAGGGCACTCCCGGAGTGTGTTGTACCCCGTGCTGTA 379

Qy   2246 rArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSe 2266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    380 CCCGGATGGGGAACAGACAGGAGCCC---AGGATGGGCTCCAGTCTCCAGGCAACAACCAG 436

Qy   2266 rGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLy 2286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 CCAGCCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAA 496

Qy   2286 sLysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGlnTyrAsnIl 2306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 GAAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATAT 556

Qy   2306 eSerGlnProGlyThrGluIlePheGluHisAlaSerThrAsnMetGlyThrLeuMetTh 2326
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 CAGCCAGCCTGGCACGGAGATCTTCAATATGCCCCATCACCGGAACA----------- 605

Qy   2326 rTyrArgSerGlnAlaValGlnHisIleAlaSerThrAsnMetGlyLeuLeuMetTh 2346
          ------------------------------------------------------------
Db    605

Qy   2346 eArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSerProLeuSerAlaAs 2366
          ------------------------------------------------------------
Db    605

Qy   2366 nAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAl 2386
          ------------------------------------------------------------
Db    605

Qy   2386 aAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyAlaLysAlaLysValSe 2406
```

```
Alignment Scores:
Pred. No.:                              Length:          846
Score:                 9.76e-51         Matches:        1448.00
Percent Similarity:    100.00%          Conservative:      0
Best Local Similarity: 100.00%          Mismatches:        0
Query Match:           10.96%           Indels:            0
DB:                                     Gaps:             13

US-09-522-753-5 (1-2517) x BU557144 (1-846)

Qy   2244  LeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGly  2263
                                          |||||||||||||||||||||||||||
Db      2  ------------------------GGTGCGCGGGCGGGGAAGGCCAAGGTCTC         61
                                          CTGCTGTACCGGATGGGAACAGACGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGC

Qy   2264  AsnThrSerGlnProProAlaPheSerLysLeuThrGluSerAsnSerAlaMetVal  2283
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62  AACACCAGCCAGCCGCCGGCCTTCTCCAAGCTGACCGAAGAGCAACTCCGCCATGGTC  121

Qy   2284  LysSerLysLysGlnIleLeuAsnLysLysLeuAsnThrHisAsnArgAsnGluProGlu  2303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122  AAGTCCAAGAAGCAAGAGATCAACAAGAGCTGAACACCCACAACCGGATGAGCCTGAA    181

Qy   2304  TyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGly  2323
                |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182  TACAATATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACGGAACAGGC   241

Qy   2324  LeuMetThrTyrArgSerGlnAlaValGlnGlnValGlnGluHisAlaSerThrAsnMetGlyLeuGlu  2343
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242  CTTATGACCTATAGAAGCCAGGCCGTGCAGGAACATGCCAGCACCAACATGGGGCTGGAG  301

Qy   2344  AlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGlnSerProProLeu  2363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302  GCCATAATTAGAAAGGCACTCATGGTAAATATGACCAGTGGGAGAGAGTCCCCGCCGCTC  361

Qy   2364  SerAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIle  2383
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362  AGCGCCAATGCTTTTAACCCTCTGAATGCCAGTGCCAGTCTGCCCGCTGCTATGCCCATA  421

Qy   2384  ThrAlaAlaAspGlyArgAspHisThrLeuThrSerProGlyGlyGlyGlyLysAla  2403
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422  ACCGCTGCTGACGGACGAGACCACACTCACCTCGCCAGGTGGCGGGGGAAGGCC       481

Qy   2404  LysValSerGlyArgProSerArgLysAlaLysSerProAlaProGlyLeuAlaSer  2423
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482  AAGGTCTCTGGCCAGACCCAGCCGAAAAGCCAAGTCCCCGGCCCTGGCCCTCATCT   541

Qy   2424  GlyAspArgProProSerSerValHisSerGluGlyGlyAspCysAsnArgArgThr  2443
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542  GGGGACCGGCCACCCTCTGTCTCCCAGTGCACTCGAGGGAGACTGCAACCGCCGGACG  601

Qy   2444  ProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPhePro  2463
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602  CCGCTCACCAACCGCGTGTGGGAGGACAGGCCCTCGTCCGCAGGTTCCACGCCATTCCCC   661

Qy   2464  TyrAsnProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGly  2483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662  TACAACCCCCTGATCATGCGCCTGCAGGCGGGTGTCATGGCTTCCCCACCCACCGGGGC  721

Qy   2484  LeuProAlaGlySerGlyProLeuAlaGlyProHisAlaTrpAspGluProLys  2503
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db    722  CTCCCCGCGGGCAGCGGGCCCCTGGCCGGCCCCCACGCCTGGGACGAGGAGCCCAAG  781

Qy   2504  ProLeuLeuCysSerGlnTyrGluThrLeuSerAspSerGlu  2517
           ||||||||||||||||||||||||
Db    782  CCACTGCTCTGCTCGCAGTACGAGACACTCTCCGACAGCGAG  823
```

```
Db    606  ------------------------GGTGCGCGGGCGGGGAAGGCCAAGGTCTC         631
Qy   2406  rGlyArgProSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyAspAr  2426
           |||||||||||||
Db    632  TGGCAGACCCAGCAGCCGANAAGGCCAAGTCCCCGGCCCTGGCATCTGGGGACCG      691

Qy   2426  gProProSerValSerValHisSerGluGlyValAspGlyAspGluThrProLeuTh  2446
Db    692  GCCACCCTCTGTCTCCCAGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCAC  751

Qy   2446  rAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAsnPr  2466
           |||||||||||||||
Db    752  CAACCGCGTGTGGGAGGACAGGCCCTCGTCCGCAGGTTCCACGCCATTCCCCTACAACCC  811

Qy   2466  oLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAla  2486
           |||||||||||||||
Db    812  CCTGATCATGCGCCTGCAGGCGGGNTGTCATGGCTTCCCCACCCACCGGGCTCCCCG  871

Qy   2486  laGlySer-GlyProLeuAlaGlyProHisHisAla---TrpAspGluProLysPro  2504
           |||   ::|
Db    872  CGGGCAGCGGGGGCCCCTGCTCGGCCCCCACCACGNCCTGGGGACCGAGGACCCAAGCCA  931

Qy   2505  Leu  2505
           |||
Db    932  CTG  934
```

```
RESULT 17
BU557144
LOCUS       BU557144              846 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AGENCOURT_10253238 NIH_MGC_109 Homo sapiens cDNA clone
            IMAGE:6584961 5', mRNA sequence.
ACCESSION   BU557144
VERSION     BU557144.1  GI:22907440
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 846)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2796 row: c  column: 09
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1..846
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6584961"
                     /tissue_type="teratocarcinoma, cell line"
                     /clone_lib="NIH_MGC 109"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: EcoRI; Site_2:
                     XhoI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

```
RESULT 18
BQ690869
LOCUS       BQ690869              984 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8034173 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6208241
            5', mRNA sequence.
ACCESSION   BQ690869
VERSION     BQ690869.1  GI:21816185
```

Qy   1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVal 1473
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACTGGCTCCAAAAAGCACGACGTA 420

Qy   1474 ArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMet 1493
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CGCTCCCTCATCGGCAGCCCCGGCGAGTTCCACCCGTGCCACCCGTGGATGTGATG 480

Qy   1494 AlaAspAlaAlaLeuGluArgAlaCysTyrGluSerLeuLysSerLeuLysSerArgProGly 1513
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGAGCCTGAAGAGCCGGCCCAGGG 540

Qy   1514 ThrAlaSer-SerSerGlyGlySerIleAlaAlaArgGlyAlaProValIleValProGluLeu 1533
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACCGGCCAGCCAGCTCGGGGGCTCCATTGCGGCGCGCCCGGTCATTGTGCCTGAGCT 600

Qy   1533 uGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaAlaProPheAlaGlyHi 1553
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGGTAAGCCGCGGCAGAGCCCCTGACCTATGAGGACCACGGGGCCACCCTTTGCCGGCCA 660

Qy   1553 sLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGluGlyLyse 1573
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CCTCCCACGAGTTCGCCGGTGACCACCGGGAGCCCACGCCGCCTGCAGGAGGAGGGCAG 720

Qy   1573 rLeuSerSerSerLysAlaSerGln-AspArgGlyLeuThrSerThrProArgGluIleA 1593
           ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   721 CCTTTCGTCCAGCAAGGCATCCCAGGACCGAAAGTCGACGTTGAGCGTTCGTAGATCG 780

Qy   1593 laLysSerProHisSerThrValProGluHisHisProHis-ProIleSer-ProTyrGl 1612
           |||||||||||||||||||||||||||||||||||||||| |||||||| |||||||
Db   781 CCAAGGTCCCGCACAGCACCGTGCCCGAGCACACCCCACACCCCATCTCGCCCTATGA 840

Qy   1612 uHisLeuLeuArgGlyVal---SerGlyValAsp-LeuTyrArgSerHis-IleProLeu 1630
           |||||||||||||| |||||| ::||| :::||| |||||||||||
Db   841 GCACCTGCTTCGCGGGGCGTGAAATGCNTGGAACCTGTATCCCCACCACCATCCCCCTG 900

Qy   1631 -AlaPheAspProThrSerIle-ProArgGlyIleProLeu 1643
            ||||| |||||||||||||||||||||||||||||
Db   901 GGCCTTCAAACCCACCTTCCATACCCCCGGCCATCCCCTCTG 941

RESULT 19
BM802749
LOCUS   BM802749   1159 bp   mRNA   linear   EST 05-MAR-2002
DEFINITION AGENCOURT_6459836 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5575037
            5', mRNA sequence.
ACCESSION   BM802749
VERSION   BM802749.1 GI:19119572
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1159)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM2324 row: k column: 06
            High quality sequence stop: 652.
FEATURES   Location/Qualifiers
     source   1..1159
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5575037"

KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 984)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2365 row: j column: 18
            High quality sequence stop: 671.
FEATURES   Location/Qualifiers
     source   1..984
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6208241"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   984
Score: 1.95e-50   Length: 984
Percent Similarity: 1442.50   Matches: 296
Best Local Similarity: 93.42% Conservative: 2
Query Match: 92.79%   Mismatches: 12
DB: 10.92%   Indels: 9
                  13   Gaps: 1

US-09-522-753-5 (1-2517) x BQ690869 (1-984)

Qy   1334 AlaIleProGluProGluArgHisSerProHisHisLeuHisLysGluGlnHisHisIleArgGly 1353
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCCATCCCGCCGGAGCCGCACAGCCCCCACCACCTCCACAAGGAGCAGCACCACCATCCGCGGG 60

Qy   1354 SerIleThrGlnGlnGlyIleLeuProArgGlySerTyrValGluAlaGlnGluAspTyrLeuArgArg 1373
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TCCATCACACAGCAGGGGATCCTCGGTTCCTACGTGGAGGCACAGGAGGACTACCTGCGTCGG 120

Qy   1374 GluAlaLysLeuLeuArgGlyLeuArgGluGlyThrProProProProSerArgAspLeu 1393
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GAGGCCAAGCTCCTAAAGCGGGAGGCGAGGGCACGCCGCCTCCGCGGGGCCCCTCACCGGACCTG 180

Qy   1394 ThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysProAlaHisGluGlyGly 1413
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACCGAGGCCTACAAGACGCAGGCCCTGGGCCCCCTGAAGCTGAAGCCGGCCCATGAGGGC 240

Qy   1414 LeuValAlaThrValLysGluAlaAlaGlyLysArgSerIleHisGluIleHisProArgGluGluLeu 1433
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTGGTGGCCACGGTGAAGGAGGCGGCCGGCTCCATCCATGAGATCCGCGGCGAGGAGCTG 300

Qy   1434 ArgHisThrProGluLysProLeuAlaProArgProLeuLysGluGlySerIleThrGln 1453
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CGGCACACGCCCGAGCTGCCCCTGCCCCGCCCGCTCAAGGAGGGCTCCATCACACAGCAG 360

```
              /tissue_type="embryonal carcinoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC 92"
              /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
              Average insert size 2.5 kb. Library enriched for
              full-length clones and constructed by Life Technologies.
              Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:               2.82e-50      Length:              1159
Score:                   1440.50      Matches:             291
Percent Similarity:      87.57%       Conservative:        12
Best Local Similarity:   84.10%       Mismatches:          20
Query Match:             10.90%       Indels:              23
DB:                                   Gaps:                2

US-09-522-753-5 (1-2517) x BM802749 (1-1159)

Qy  1541 LeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProVal 1560
         ||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db     2 CTGACCTATGAGGACCACGGGGCACCCTTGCCGGCCACCTCCCACAGAGGTTCGCCCGTG 61

Qy  1561 ThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSer 1580
         |||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 ACCATGCGGGAGCCCACGCCGCGCCTGCAGGAGGGCAGCCCTTTCGTCCAGCAAGGCATCC 121

Qy  1581 GlnAspArgGlyLeuThrSerThrProArgGluIleLysSerProHisSerThrVal 1600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CAGGACCGAAAGCTGACGTCGACGCCTCGTGAGATCGCCAAGTCCCCACACAGCACCGTG 181

Qy  1601 ProGluHisHisProHisProIleSerSerProTyrGluHisLeuLeuArgGlyValSerGly 1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 CCCGAGCACCACCCACACCCCATCTCGCCCTATGAGCACCTGCTTCGGGGCGTGAGTGGC 241

Qy  1621 ValAspLeuTyrArgSerHisIleProLeuAlaPheAspAspProThrSerIleProArgGly 1640
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   242 GTGGACCTGTATCGCAGCCACATCCCCCTGGCCTTCGACGACCCCACCTCCATACCCCGGC 301

Qy  1641 IleProLeuAspAlaAlaAlaAlaTyrTyrProArgHisLeuAlaProAsnProThr 1660
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 ATCCCTCTGGACGCAGCCGCTGCCTACTACCCTCGCCCCGACACTCGGCCCCCAACCCCACC 361

Qy  1661 TyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGlu 1680
         |||||||||||||||||||||||||||||||||||||||||||||
Db   362 TACCCGCACCTGTACCCACCCTACCTCATCCGCGGCTACCCGGACACGGCGGCGCTGGAG 421

Qy  1681 AsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnMetHisHisAsnThrAla 1700
         |||||||||||||||||||||||||||||||||||||||||||||
Db   422 AACCGGCAGACCATCATCAATGACTACATCACCTCGCAGATGCACCACAACACGGC 481

Qy  1701 ThrAlaMetAlaGlnAsnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerLeu 1720
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 ACGGCCATGGCCCAGAACCGCGCCGATATGCTGAGGGGCCTCTCGCCCCGGGAGTCCTCGCTG 541

Qy  1721 AlaLeuAsnTyrAlaAlaGlyTyrLeuIleIleAspLeuAspLeuAsnValProHisGluLeu 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 GCACTCAACTACGCTGCGGGTTCCCGAGCGATCATCGACCTGGACCTGAATGTGCCACACTG 601

Qy  1741 ProValLeuValProProGlyThrProAlaThrAlaMetAspArgGlyLeuAlaTyr 1760
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 CCTGTGCTGGTGCCCCCGACACCAGGACACCCGGCCACCGCCATGGACCGCGGCCTTGCCTAC 661

Qy  1761 LeuProThrAlaProGln-ProPheSerSerArgHisSerSerProLeuSerProGl 1780
         ||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   662 CTCCCACCGCGCCCCAGCCCCTTCAGCAGCCGCCACAGCAGCTCCCCACTCTCCCCAGG 721

Qy  1780 yGlyProThrHisLeuThrLysProThrThrSerSerGluArgGluArgAspAr 1800
         ||||||||||||||||||||||||||||||||||||||||  |::  ||
Db   722 AGTCCAACAACTTGACAAAACCAACCAACACGTCCTCCTGAACGGGAGCGAGCGAAACCG 781

Qy  1800 gAspArgGlu-ArgArgAspArg-GluArgGluLysSerIleLeuThrSerThr-Th 1819
         |:::||||  ||||||| |::||| ::||||  ::||||||||||
Db   782 GAATCGAGAGAGGACGGACCGGATCGGGACCGGGAAAGTCCTTCTTCACGTCCACCAAC 841

Qy  1819 rThrValGluHisAlaProIleTrp-ArgProGly-ThrGluGlnSerSerGlySerSer 1838
         ||||||     :::||||||| ||:|| |||||||
Db   842 AACGGTGGGACCCGGACCACCCTCTGGAAACCTGGTAACCGAAGAAAAATACCGGACGCAA 901

Qy  1839 GlySerSerGlyGlyGlyGlySer------------------SerSerArgPro 1851
         |||:||  |:::                          :::::||||||
Db   902 GGGAACACAACGAGGGGGGGTGGTACACCTTCACAGAATCATACCTACGCTGCTCGTCCT 961

Qy  1852 ----------------------AlaSerHisSerHisAlaHisGlnHisSer 1861
                               |||||:::|||||
Db   962 ATTGCTGTAGAGCACGACTGGATACGTTCCGTCACATACCTGTAATTATATCATAGT 1021

Qy  1862 ProIle 1863
         ||||||
Db  1022 CCGATC 1027
```

```
RESULT 20
BM910785

LOCUS       BM910785              1022 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6617197 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454337
            5', mRNA sequence.
ACCESSION   BM910785
VERSION     BM910785.1  GI:19361164
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1022)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1949 row: f  column: 02
            High quality sequence start: 12
            High quality sequence stop: 554.
FEATURES             Location/Qualifiers
     source          1..1022
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454337"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:               1.74e-49      Length:              1022
Score:                   1420.00      Matches:             289
Percent Similarity:      86.74%       Conservative:        12
Best Local Similarity:   83.29%       Mismatches:          26
Query Match:             10.75%       Indels:              21
DB:                                   Gaps:                4
```

Db   993 CTCTGCCAAACCCCC 1007

RESULT 21
Db   BM558844
LOCUS   BM558844   1030 bp   mRNA   linear   EST 20-FEB-2002
DEFINITION   AGENCOURT_6592838 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5476852
                5', mRNA sequence.
ACCESSION   BM558844
VERSION   BM558844.1 GI:18801929
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1030)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: DCTD/DTP
                cDNA Library Preparation: Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Agencourt Bioscience Corporation
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLCM1993 row: p column: 05
                High quality sequence stop: 618.
FEATURES   Location/Qualifiers
     source   1..1030
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5476852"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_41"
                     /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 1030
Score: 6.25e-48   Matches: 288
Percent Similarity: 85.76%   Conservative: 7
Best Local Similarity: 83.72%   Mismatches: 29
Query Match: 10.46%   Indels: 21
DB: 12   Gaps: 5

US-09-522-753-5 (1-2517) x BM558844 (1-1030)

Qy   1457 LeuLysTyrAspThrGlyValAlaSerThrThr-GlySerLysHisAspValArgSerIle 1476
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   2 CTCAAGTACGACACAGGCGCCGTCCACCACTGGCTCGGCTCCAAAAGCACGACGTACGTCCCT 61

Qy   1476 uIleGlySerProGlyArgThrPheProGlyValHisProLeuHisProLeuMetAlaAspAl 1496
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 CATCGGCAGCCCCGGCCGAGCGTTCCACCCTGCACCCTGATGTGATGGCCGACGC 121

Qy   1496 aArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSe 1516
          |||||||||||||||||||||||| |||||||||||||||||
Db   122 CCGGGCACTGGAACGTGCCTGCTACGAGGAGAGCCTGAAGAGCCGGCCAGGGACCGCCAG 181

Qy   1516 rSerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluGluGlyLysPr 1536
          |||||||||||||||||||||||||||||| |||||||||||||||
Db   182 CAGCTCGGGGGGCTCCATTGCGCGCGGCGCCCCGTCATTGTGCCTGAGGAGGGCAAGCC 241

US-09-522-753-5 (1-2517) x BM910785 (1-1022)

Qy   2151 HisProGlnGlnLeuSerAlaProAlaProLeuProTyrSerPheProGlyAlaSer 2170
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13 CACCCACAGCAGCTCAGGCGCACCCCTGCCCGCCCCCCCTCACTCCTTCCCTGGGGCCAGC 72

Qy   2171 CysProValLeuAspLeuArgArgProSerArgAspLeuTyrLeuProLeuProProAspHis 2190
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 TGCCCCGTCCTGGACCTCCGCCGCCCCCACCCAGTGACCTTCACCTCCCGCGCCCCCGACCAT 132

Qy   2191 GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgGlySerProGluProAsn 2210
          |||||||||||||||||||||||||||||||||||||||||||||
Db   133 GGTGCCCCCGGCCCGCGGTCCCCCCCACAGCGAAGGGGCAACAGAGTCTCCAGAGCCAAAC 192

Qy   2211 LysThrSerValLeuGlyGlyGlyGluAspGlyIleGluGluProValSerProGlyGluGly 2230
          |||||||||||||||||||||||||||||||||||||||||||||
Db   193 AAGACGTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGGGC 252

Qy   2231 MetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlu 2250
          ||||||||||||||||||||||||||||||||||||||||||||
Db   253 ATGACGGAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGGATGGGGAA 312

Qy   2251 GlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla 2270
          |||||||||||||||||||||||||||||||||||||||||||||
Db   313 CAGACGGAGCCCAGCCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCC 372

Qy   2271 PhePheSerLeuThrGluSerAsnSerAlaMetValLysSerLysValSerGlnGlnIle 2290
          ||||||||||||||||||||||||||||||||||||||||||||
Db   373 TTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGGTCAGTCAGCAGATC 432

Qy   2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnInProGly 2310
          |||||||||||||||||||||||||||||||||||||||||||||
Db   433 AACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCCTGGG 492

Qy   2311 ThrGluIlePheAsnMetProAlaIleThrGlyTyrGlyLeuMetThrTyrArgSerGln 2330
          ||||||||||||||||||||||||||||||||||||||||||||
Db   493 ACGGAGATCTTCAATATGCCCGCCATCACCGGACACAGGCCTTATGACCTATAGAAGCCAG 552

Qy   2331 AlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluLeuAlaIleIleLeuArgLysAlaLeu 2350
          ||||||||||||||||||||||||||||||||||||||||||||
Db   553 GCGGTGCAGGAACATGCCAGCACCAACATGGGCCTGGAGCTGGCCATAATTAGAAAGGCACTC 612

Qy   2351 MetGlyLysTyrArgAspGlnInTrpGluGluSerProLeuSerAlaAsnAlaPheAsnAspPro 2370
          |||||||||||||||||||||||||||||||||||||||||||||
Db   613 ATGGGTAAATATGACCAGTGGGAAGAGTCCCCGCTCAGCGCCAATGCTTTTAACCCT 672

Qy   2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaLaAspAspGlyArgSer 2390
          ||||||||||||||||||||||||||||||||||||||||||
Db   673 CTGAATGCCAGTGCCAGCCTGCCCGCTGCTATGCCCATAACCGTGCTGACGGACGGAGT 732

Qy   2391 AspHisThrLeuThrSerProGlyGlyGlyLysGlyLysValSerAlaValSerGlyArg-ProSe 2410
          |||||||||||||||||||||| |::: ||:::
Db   853 TGTCTCCTCAGGGCCCTCCGAGGGAAAAATGGCACCCCCGGAA-ACCCCGTTCACCCAA 911

Qy   2410 rSerArgLysAlaLysSerProAla-ProGlyLeuAlaSerGlyAsp-ArgProProSer 2429
          :::|||||:::|||||||||||||| |::: |||:::
Db   793 AACCTAAAAAGCCAGGTCCCCGGCTCCCATCTGGGGAACCCGGCCACCCTCC 852

Qy   2430 ------ValSerSerValHisSerGluGlyValAspCysAsnArgArgThrProLeuThrAsn 2447
          || :::   :::::::
Db   853 TGTCTCCTCAGGGCCCTCCGAGGGAAAAATGGCACCCCCGGAA-ACCCCGTTCACCCAA 911

Qy   2448 --ArgValTrpGluAspArgProSerSer---AlaGlySerThrProPheProTyrAsn 2465
          ||| :::   :::|||||:::
Db   912 CCGGGTGGGGAAAAAACAGCCCTCCGCCCCCCAGGCTCCAC------ 959

Qy   2466 ProLeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProProGlyLeuProPro 2485
          ||||||||||||||||||||||||| :::|||
Db   960 ------------------------------GTTTCCCCTCCAACCCCCTTCAAACATTCCG 992

Qy   2486 AlaGlySerGlyPro 2490
          |||

COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapps-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2007 row: c   column: 16
            High quality sequence stop: 661.

FEATURES   Location/Qualifiers
     source   1..1010
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5481927"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_41"
                     /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   8.95e-48   Length: 1010
Score: 1378.50   Matches: 281
Percent Similarity: 85.50%   Conservative: 8
Best Local Similarity: 83.14%   Mismatches: 31
Query Match: 10.43%   Indels: 19
DB: 12   Gaps: 3

US-09-522-753-5 (1-2517) x BM915686 (1-1010)

```
Qy     1457 LeuLysTyrAspThrAlaSerThrGlyProPheAlaProPheAlaGlyHisAspValArgSerLeu 1476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 CTCAAGTACGACACACCGGCGCGTCCACCACTGGCTCCAAAAGCACGACGTACGGTCCCTC 60

Qy     1477 IleGlySerProGlyArgThrPhePheProProValHisProLeuAspValMetAlaAspAla 1496
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 ATCGGCAGCCCGGCCGGACGTTCCACCCCTGCACCCGTGATGTGATGGCCGACGCC 120

Qy     1497 ArgAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgGlyThrAlaSer 1516
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 CGGGCACTGGAACGTGCCTACGAGAGAGCCTGAAGAGAGCCGGCCAGGGAGCCGCCAGC 180

Qy     1517 SerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysPro 1536
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGCTCGGGGGGCTCCATTGCGCGGGCGGCCCCGGTCATTGTGCCTGAGCTGGGCAAGCCG 240

Qy     1537 ArgGlnSerProLeuThrTyrGluAspHisGlyValAspProPheAlaGlyHisLeuProArg 1556
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CGGCAGAGCCCCCTAACCTATGAGGACCACGGGCACCCTTTGCCGGCCACCTCCACGA 300

Qy     1557 GlySerProValThrMetArgGluProThrArgLeuGlnGluGlySerLeuSerSer 1576
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GGTTCGCCGTGACCACGCGGCAGCCCACGCGCCTGCAGGAGGGGCAGCGCCTTTCGTCC 360

Qy     1577 SerLysAlaSerGlnAspArgLysLeuThrSerProArgGlyIleAlaAlaLysSerPro 1596
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 AGCAAGGCATCCCAGGACCGAAAGCTGACGTCGCCGCGTGGGATCGCGAAGTCCCCG 420

Qy     1597 HisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuArg 1616
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 CACAGCACCGTGCCCGAGCACCACCCACACCCATCTCGCCCTATGAGCACCTGCTTCGG 480

Qy     1617 GlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSer 1636
```
```
Qy     1536 oArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaProPheAlaGlyHisLeuProAr 1556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      242 GCCGCAGAGCCCCCTAACCTATGAGGACCACGGGCACCCTTTGCCGGCCACCTCCACG 301

Qy     1556 gGlySerProValThrMetArgGluProThrArgLeuGlnGluGlySerLeuSerSer 1576
            ||||||||||  |||||||||||||||||||||||||||||||||||||||||||
Db      302 AGGTTCGCCGTGACCACGCGGCAGCCCACGCGCCTGCAGGAGGGGCAGCGCCTTTCGTC 361

Qy     1576 rSerLysAlaSerGlnAspArgLysLeuThrSerProArgGlyIleAlaAlaLysSerPr 1596
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      362 CAGCAAGGCATCCCAGGACCGAAAGCTGACGTCGCCGCGTGGGATCGCGAAGTCCCCC 421

Qy     1596 oHisSerThrValProGluHisHisProHisProIleSerProTyrGluHisLeuLeuAr 1616
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      422 GCACAGCACCGTGCCCGAGCACCACCCACACCCATCTCGCCCTATGAGCACCTGCTTCG 481

Qy     1616 gGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSe 1636
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      482 GGGCGTGAGTGGGGTGGACCTGTATCGGCAGCCACATCCCCTGCCTTCGACCCCACCTC 541

Qy     1636 rIleProArgGlyIleGluProLeuAspAlaAlaAlaAlaAlaThrTyrTyrLeuProArgHisLeuAl 1656
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      542 CATACCCCGCGGCATCCCCTCTGGACGCAGCCGCTGCCTACTACCTGCCCCGACACCTGGC 601

Qy     1656 aProAsnProThrTyrProHisLeuTyrProTyrLeuIleLeuArgGlyTyrProAspTh 1676
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      602 CCCCAACCCCACCTACCCGCACCTGTACCCCTACCTCATCCGGGCTACCCCGACAC 661

Qy     1676 rAlaAlaLeuGluAsnThrAspArgGlnThrIleIleGluAsnAspTyrIleThrSerGlnGlnMetHi 1696
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      662 GGCGGCACTGGAGAACACCGACAGACAGACCATCATCGAGAATGACTACATCACCTCGCAGCAGATGCA 721

Qy     1696 sHisAsnThrAlaThrAlaMetAlaGlnAsnArgAlaAspMet-LeuArgGlyLeuSerProA 1716
            ||||||   |||||||||||||||||||||     |||||||||||||||||||
Db      722 CCACAACGGCGCCACCGCCATGGCCCAGAGCCTGATATGCTTGAGGGGCCTCTCGCCCC 781

Qy     1716 rg-GluSerSerLeuAlaLeuAsnTyrAla------AlaGlyProArgGlyIleIleIleAsp 1733
            ||   ||||||||||||||||||||         |||
Db      782 GGCGGAGACCTCGCTGGCACTCAACTACGCGTGGCGCGGGGTCCCCCAGGAAGGTC------ 835

Qy     1734 LeuSerGlnValProHisLeu---------------ProValLeuProThrProProGly 1749
            ||||||||||||                    |||||||||:::|||||||||||
Db      836 --------CCACCCACTTGGACAAAACCAACCACCCCT--CTCCTTCCCCCCTCCCCAGCG 884

Qy     1750 ThrProAlaThrAlaMetAspArgLeuAlaTyrLeuTyrLeuPro-ThrAlaProGlnPro------ 1767
            |||                 |||             ::|||
Db      885 GGGGAGCCGAAACCCGGGGATTCAACACGGGCAACCCGGCATCGCGGAAGCCGGGACA 944

Qy     1768 ----PheSerSerArgHisSerSerSerProLeuSerProGly------GlyProThrHi 1784
            ||:::||||            |||:::||
Db      945 CAATATTTCTTCGTCCATGCTCCCCCCCAGATCGGTCGGCCAGCCCCGGCCGCCCCA 1004

Qy     1784 sLeu 1785
            ||||
Db     1005 CCTC 1008
```

RESULT 22
BM915686
LOCUS   BM915686   1010 bp   mRNA   linear   EST 12-MAR-2002
DEFINITION AGENCOURT_6639688 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5481927
            5', mRNA sequence.
ACCESSION   BM915686
VERSION   BM915686.1 GI:19366065
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1010)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)

Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   1.19e-47
Score: 1374.00   Length: 887
Percent Similarity: 85.21%   Matches: 265
Best Local Similarity: 85.21%   Conservative: 0
Query Match: 13   Mismatches: 0
DB: Indels: 46
                                           Gaps: 1

US-09-522-753-5 (1-2517) x BU542258 (1-887)

```
Qy   2207 ProGluProAsnLysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSer 2226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCAGAGCCAAACAACAACGTCGGTCTTGGGTGGTGGTGGAGGACGGTATTGAACCTGTCC 60

Qy   2227 ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCACCCGGAGGGCATGACGAGACCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTAC 120

Qy   2247 ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGGATGGGGAACACGACGACGAGCCCAGCCGCAGGAGGGCTCCAAGTCTCCAGGCAACACCAGC 180

Qy   2267 GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys 2286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGCCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAG 240

Qy   2287 LysGlnGluIleAsnLysLysLysAsnThrHisAsnArgAsnGluProGluTyrAsnIle 2306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGCAAGAGATCAACAAGAAGAAGAAGAACACCCACAACCGAAATGAGCCTGAATACAATATC 300

Qy   2307 SerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThr 2326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACC 360

Qy   2327 TyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIle 2346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TATAGAGACCAGGCGGTGCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATT 420

Qy   2347 ArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSerProLeuSerProAlaAsn 2366
          ||||||||||||||||||||
Db    421 AGAAAGGCACTCATG                                             435

Qy   2367 AlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAla 2386
          ---------------------------------------------------------
Db    436 ----------------------------------------------------------- 435

Qy   2387 AspGlyProSerArgLysAlaAlaProGlyLeuAlaProGlyLysAlaLysLysValSer 2406
                                             |||||||||||||||||||||||
Db    436 --------------------------GGTGGCGGCGGGAAGGCCAAGGTCTCT 462

Qy   2407 GlyArgProSerArgLysAlaLysSerProAlaSerGlyAlaProGlyLeuAlaSerGlyAspArg 2426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 GGCAGACCCAGCCAGCCGAAAAGCCAAGTCCCCGGCCTCGGGCATCTGGGGACCGG 522

Qy   2427 ProProSerValSerSerValHisSerGluLysCysAsnArgArgThrProLeuThr 2446
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 CCACCCTCTGTCTCCTCAGTGCACTCGGAGGGAGACTGCAACCGCGGACGCGGCTCACC 582

Qy   2447 AsnArgValTrpGluArgProSerSerAlaGlySerThrProPheProTyrAsnPro 2466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    583 AACCCGCGTGTGGGAGGACAGGCCCCTTCCGTCCGCAGTTCCACGCATTCCCTACAACCCC 642

Qy   2467 LeuIleMetArgLeuGlnAlaGlyValMetAlaSerProProGlyLeuProAla 2486
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    643 CTGATCATGCGGCTCCAGGCGGGTGTCATGGCTTCCCCACCCCACCCCCACCCGGGGCCTTCCCCGCG 702
```

```
Db    481 GGCGTGAGTGGGTGGCACCTGTATCGCAGCCACATCCCCCTGGCCTTCGACCCCCACCTCC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1637 IleProArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAla 1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATACCCGCGGCATCCCTCTGGACGCAGCGCGCTGCCTACTACCTGCCCCGACACCTGGCC 600

Qy   1657 ProAsnProThrTyrProHisLeuTyrProProTyrIleLeuArgGlyTyrProAspThr 1676
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCCAACCCCACCTACCCGCACCTGTACCCGCCCTACCCCATCCTGCGCGGCTACCCCGACACG 660

Qy   1677 AlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnHisGlnMetHis 1696
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GCGGCGCTGGAGAACCGCCAGACCATCATCAATGACTACATCACCTCGCAGCAGATGCAC 720

Qy   1697 HisAsnThrAlaThrAlaMetAlaGlnArgAlaAsp-MetLeuArgGlyLeuSerProA 1716
          ||||||| ||||| :::|||
Db    721 CACAACACGGCTCCCGCCATGGCCCAGCTGAGCTAATATTGCTGAGGGGCCTCTCCGCCC 780

Qy   1716 rgGluSerSerLeuAlaLeuAsnTyrAlaAlaGly---ProArgGlyIleIleAspLeuS 1735
          ||||||||||||||||||||  :::
Db    781 GCGAGTCCTCGCTGGCACTCCACTACCCCGCGGGGGTCCCCAGGGAGGTCAAAAACCCTTG 840

Qy   1735 erGlnValProHisLeuProValLeu----------ValProProThrPro----- 1748
          ::|||  |||
Db    841 GCCAAA--CCAACCCGCCGGTCCTTCGCCCAACGGCAGTGAGCCCCGGAATCCAAACG 898

Qy   1749 ----------GlyThrProAlaThrAlaMetAspArgArgLeuAlaT 1760
          ::|::|||
Db    899 CGTACCCGGATCCGGCAGGGGAAAAGGCACTGTTTACGTCCCACCTGAAAGGTGAAAAC 958

Qy   1760 yrLeuProThrAlaProGlnProPheSerSerArgHisSerSerSerPro 1776
          :: ||| ::: :::|||
Db    959 ATGGGGCCCCTCCTTGTAAACCGTGAAACATGAGGAGACTCACGGGTCC 1008
```

RESULT 23
BU542258
LOCUS   BU542258   887 bp   mRNA   linear   EST 13-SEP-2002
DEFINITION AGENCOURT_10253089 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6574079
            5', mRNA sequence.
ACCESSION   BU542258
VERSION   BU542258.1   GI:22852741
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 887)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2768 row: m column: 23
            High quality sequence stop: 710.
FEATURES   Location/Qualifiers
     source   1..887
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6574079"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_40"
                     /note="Organ: Prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.

```
Qy   952  ArgAlaAsnAlaSerProGlnLysProLeuAspLeuLysGlnLeuLysGlnArgAlaAla  971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CGGGGCCAATGCCTCACCCCAGAGCCACTGGACCTGAAGCAGCTGAAGCAGCGAGCGGCT  240

Qy   972  AlaIleProIleGlnValThrLysValHisGluProProArgGluAspAlaAlaPro  991
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  GCCATCCCCCCATCCAGTCACCAAAGTCCATGAGCCCCCGGGAGGACGCAGCTCCC  300

Qy   992  ThrLysProAlaProProAlaProProGlnAsnLeuGlnProGlnSerAspArgAla  1011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  ACCAAGCCAGCTCCCCACCGCCAAAACCTGCAGCCGGAGAGCGAGCGCGCC  360

Qy   1012 ProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAlaProProAlaAsp  1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  CCTCAGCAGCTGGCAGCAGCCCGGGGCAAGAGCAGGAGCCCCGGCACCCCCCGCCGAC  420

Qy   1032 ----------------AlaPheAlaAlaGluAlaGlnLysLeuPro  1043
                          ||||||||||||||||||||||||||||
Db   421  AAGGAGGACGAGAAGCCTGTGTTCTCCCAGCCTTCGCAGCCCAGAAGCTGCCT  480

Qy   1044 GlyAspProCysTrpThrSerGlyLeuProPheProValProAlaProGluValIle  1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GGGGACCCCCCTGCTGGACTTCCGGGCCTGCCCTTCCCCGTGCCCCCCGTGAGGTGATC  540

Qy   1064 LysAlaSerProHisAlaProAlaAspProSerAlaPheSerTyrAlaProProGlyHisPro  1083
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  AAGGCCTCCCCCATGCCCCCGACCCCTCAGCCTTCTCCACGCTCCCACCTGGTCACCCA  600

Qy   1084 LeuProLeuGlyLeuHisAspThrAlaArgProValLeuProArgProProThrIleSer  1103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  CTGCCCCTGGGCCTTCCATGACACTGCCCGGCCCGTCCTGCCGCGCCCCACCATCTCC  660

Qy   1104 AsnProProLeuIleSerSerArgAlaLysHisProSer-ValLeuUArgGlnIleGl  1123
          |||||||||||||||||||||||||||||||||||||  ||||||||||||
Db   661  AACCCGCCTCCCCTCATCTCCTGCCAAGCACCCCAGCGNTCCTCGAGAGGCAAATAGG  720

Qy   1123 yAlaIleSerGln-GlyMetSerValGlnLeuHisValProTyrSerGluHisAlaLys-  1142
          ||||||||||||| ||||||||||||||||||||||||||||||||||    :::
Db   721  TGCCATCTCCCAAGGGAATGTCGGTCCAGCTCCAGCTCCGTACTCAGCAGGG  780

Qy   1143 AlaProValGlyProValThrMet---GlyLeuProLeuProMetAspProLysLysLeu  1161
          ||||||||||||||||||||||   |||||||||||||||||::::
Db   781  GCCCCGGTGGGCCCCGTGTCACCATGGGGCGCTGCCCCTGCCCATGGAACCCCAAAAGCT  840

Qy   1162 ---AlaProPheSerGlyValLys-GlnGluGlnLeuSerPro----ArgGlnGlnAla--  1178
             :::|||||||||||||  :::||||||||||||     ||||||||||||
Db   841  GGGACCCCTTCAACGGAGTGAAACCAGGAACCAGCTGGTCCCCACCCGGGGCCCAGGCCTG  900

Qy   1179 GlyProProGluSerLeu------------GlyValProThrAlaGlnAlaSerValLeuArg  1196
          |||||||||||           |||              :::|||  :::
Db   901  GGGGCACCGGAAAAACCCTGGGGGGGGCCCCACACGCCCCGAAAGCC-------GTCCCG  954

Qy   1197 GlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGly-------IlePro  1213
          |||                       :::||||::::|||:::
Db   955  GGGCTGGAGAAGGGACAGCTTCCGGGGCCTTCAATTCCGGGGGGGAAAGCCTTCCCCC  1014

Qy   1214 SerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThr  1230
          ||||||||||||||||||||::::::
Db   1015 CAAAGGGATTTTCCCAGGAACACGGGGGCCCCCTCCGAAAGAGCGCCACT  1065
```

RESULT 25
BX368971

```
Qy   2487 GlySerGlyLeuAlaGlyProHisHisAlaTrpAspGluProLysProLeuLeu  2506
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   703  GGCAGCGGCGGCCCTGCTGGCCCCCCACACGCCTGGGACGAGAGGCCCAAGCCCACTGCTC  762

Qy   2507 CysSerGlnTyrGluThrLeuSerAspSerGlu  2517
          ||||||||||||||||||||||
Db   763  TGCTCGCAGTACGAGACACTCTCCGACACGCGAG  795
```

RESULT 24
BM472005

ORIGIN

Alignment Scores:
```
Pred. No.:                          1..1067
Score:              2.59e-47        Length:        1067
                    1368.00         Matches:        289
Percent Similarity: 82.83%          Conservative:    10
Best Local Similarity: 80.06%       Mismatches:      38
Query Match:        10.35%          Indels:          24
DB:                 12              Gaps:             7
```

```
US-09-522-753-5 (1-2517) x BM472005 (1-1067)

Qy   892  LysAlaGluLysLysGluGlyGlyLySerGlyArgAlaThrThrAlaLysSerGlyAla  911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1    AAGGCAGAAGAAGAAGGAGGGCGGAGCCGGAGGCCACCACCAGCCAAGAGCTCGGCGCGC  60

Qy   912  ProGlnAspSerAspSerAlaThrCysSerAspSerAlaAspGluValAlaGluGly  931
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   CCCCAGGACAGCGACCTCCAGTGCCACCTGCAGTGCAGACGAGGTGGATGAGGCCGAGGGC  120

Qy   932  GlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrProThrGlyAspPro  951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GGGCGACAAGAACCGGCTGCTGTCCCCAAGGCCCCAGCCTCCTCACCCCCGACTGGCGACCCC  180
```

```
AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of
             Invitrogen. This sequence belongs to sequence cluster 6282.r For
             more information about this cluster, see
             http://www.genoscope.cns.fr/
             cgi-bin/cluster.cgi?seq=CS0BAE0292G02_AE02704_1&cluster=6282.r.
             Contact : Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
             Faraday Avenue Genoscope sequence ID : CS0BAE0292G02_AE02704_1.
FEATURES     Location/Qualifiers
     source  1..868
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="CS0DL010YH07"
             /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
             /cell_line="RAMOS CELL LINE"
             /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
             25-NORMALIZED"
             /note="1st strand cDNA was primed with a NotI-oligo(dT)
             primer. Five prime end enriched, double-strand cDNA was
             digested with the Not I and cloned into the Not I and EcoR V
             sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:            2.38e-47
Score:               1366.50
Percent Similarity:  92.81%
Best Local Similarity: 91.78%
Query Match:         10.34%
DB:                  13
```

```
                           Length:         868
                           Matches:        268
                           Conservative:   3
                           Mismatches:     13
                           Indels:         9
                           Gaps:           3
```

US-09-522-753-5 (1-2517) x BX368971 (1-868)

```
Qy  1590 ArgGluIleAlaLysSerProHisSerThrValProGluHisProHisProIleSer 1609
              |||:::||||           ||||||||||||||||||||||
Db     2 CGTGGGATCGCACAGTCCCCGAC-AGCACCGTGCCCGAGCACCACCCCATCTCG   60

Qy  1610 ProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIlePro 1629
              |||||||||||||||||||||||||||||||||||||
Db    61 CCCTATGAGCACCTGCTTCGGGGCGTGAGTGGCGTGGACCTGTATCGCAGCCACATCCCC  120

Qy  1630 LeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAla-AlaAlaAlaTy 1649
              |||||||||||                ||||||||||||||||
Db   121 CTGGCCTTCGACCCCACCTCCATACCCCGCGGCATCCCTCTGGACGCCAGCCGCTGCCTA  180

Qy  1649 rTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrIle 1669
              ||||||||||||||||||||||||||||||||||||
Db   181 CTACCTGCCCCGACACCTGGCCCCCAACCCCACCTACCCGCACCTGTACCCACCCTACCT  240

Qy  1669 uIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTy 1689
              ||||||||||||||||||||||||||||||||||||
Db   241 CATCCGCGGCTACCCCGACACCGCCGCGCTGGAGAACCGGCAGACCATCAATGACTA    300

Qy  1689 rIleThrSerGlnIleMetHisHisAsnThrAlaThrAlaMetAlaGlnAlaAspAspMe 1709
              |||||||||||||||||||||||||||||||||||
Db   301 CATCACCTCGCAGATGCACCACACACGCGGCCACATGGCCCAGCGAGCTGATAT  360

Qy  1709 tLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProAr 1729
              |||||||||||||||||||||||||||||||||||
Db   361 GCTGAGGGGCCCTCCGCCCCGCGAGTCTTCGCTGGCACTCAACTACGCTGCGGGTCCCCG  420

Qy  1729 gGlyIleIleAspLeuSerGlnValProHisLeuProValLeuValProProThrProGl 1749
              ||||||||||||||||||||||||||||||||||
Db   421 AGGCATCATCGACCTGTCCCAAGTGCCACACCTGCCCGTGCTCGTGCCCCCGACACCAGG  480
```

```
Qy  1749 yThrProAlaThrAlaMetAspArgLeuAlaTyrLeuGlnProThrAlaProGlnProPheSe 1769
              |||||||||||||||||||||||||||||||||||||
Db   481 CACCCCAGCCACCGCCATGGACCGGCTTGCTACCTCCCACCGCGCCCAGCCCTTCAG  540

Qy  1769 rSerArgHisSerSerProLeuSerProGlyGlyProThrHisLeuThrLysProTh 1789
              |||||||||||||||||||||||||||||||||||
Db   541 CAGCCGCCACAGCAGCTCCCCCTCTCCCCAGGAGTCCAACACACTTGACAAAACCAAC  600

Qy  1789 rThrThrSerSerGluArgGluArgArgArgGluArgGluArgAspArgArgArgl 1809
              |||||||||||||||||||||||||||||||||||
Db   601 CACCACGTCCTCGTCCGAGCGGGAGCGGGAGCGCGAGAGCGGGACCGGGATCGGGA  660

Qy  1809 uArgGluLysSerIleLeuThrSerThrThrThrValGluHisAlaProIleTrpArgPr 1829
              ||||||||||||||||||||||||||||||||||
Db   661 GCGGGAAAAGTCCATCTCACGTCCACCACGACGGTGGAGCACGCACCCATCTGGAGACC  720

Qy  1829 oGlyThrGlnGlnSerSerGlySerSer------GlyGlyGlyGlySer 1847
              |||||||||||||:::                 ||||||:
Db   721 TGGTACAGAGCANAGCAGCAGCGGCAGCACGCGGCAGCACGGGGGGGTTGGGGGNNNNN  780

Qy  1847 rSerSerArgProAlaSerHisSerHisAlaHisGln--------HisSerProIleSe 1864
              |||||||||||||||||||||||||:::
Db   781 NNNNNNNNNCCCNNCTTCCCACTCCCATGCCCACCAAGCACTTGCCCATTTTCCCTTGGG  840

Qy  1864 rProArgThrGlnAspAlaLeuGlnGlnArgPro 1875
              ||||||||||||||||
Db   841 ACCCAGG------AAGGCCCTCCAACAGAGACCA  868
```

```
RESULT 26
BU184403
LOCUS        BU184403               923 bp    mRNA    linear    EST 04-SEP-2002
DEFINITION   AGENCOURT_7933306 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6172730
             5', mRNA sequence.
ACCESSION    BU184403
VERSION      BU184403.1  GI:22698387
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 923)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
  COMMENT    Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM13543  row: c  column: 03
             High quality sequence stop: 625.
FEATURES     Location/Qualifiers
     source  1..923
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:6172730"
             /tissue_type="retinoblastoma"
             /lab_host="DH10B (phage-resistant)"
             /clone_lib="NIH_MGC_67"
             /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
             Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
             Average insert size 1.75 kb. Library constructed by Life
             Technologies."
ORIGIN

Alignment Scores:
Pred. No.:           1.42e-46          Length:          923
Score:               1348.50           Matches:         280
```

```
Percent Similarity:      92.72%        Conservative:        0
Best Local Similarity:   92.72%        Mismatches:          7
Query Match:             10.20%        Indels:             15
DB:                      13            Gaps:                3

US-09-522-753-5 (1-2517) x BU184403 (1-923)

Qy   828  GluGluGluThrAlaAlaAlaProProValGluGluGlyGluGlnLysProProAla  847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2  GAGGAGGAGACCGCAGCGGCGCCCCCAGTGGAGGAGGGGGAGGAGCAGAAGCCCCCGCG   61

Qy   848  AlaGluGluLeuAlaValAspThrGlyLysAlaAspGluProValLysSerGluCysThr  867
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62  GCTGGAGGAGCTGGCAGTGACACAGGGAGGAAGGCCCGTCAAGAGCGAGTGCACG  121

Qy   868  GluGluAlaGluGluGluProAlaLysGlyLysAspAlaGluAlaAlaGluAlaThrAla  887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122  GAGGAAGCCGAGGAGGGGCCGGCCAAGGGCAAGGACGCGGAGGCGCTGAGGCCACGGCC  181

Qy   888  GluGlyAlaLeuLysGluLysSerGlyGlySerGlyArgAlaThrAlaLys  907
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182  GAGGGCGCGCTCAAGGCACGAGAAGGAGGGGCGGAGCGGGCCACCAGCCAAG  241

Qy   908  SerSerGlyAlaProGlnAspSerAspSerAlaThrCysSerAlaAspGluValAsp  927
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242  AGTCTGGGGCCCCCCAGGACACGCGACTCCAGTGCCACCTGCAGTGCAGAGGTGGAT  301

Qy   928  GluAlaGluGluGlyGlyAspLysAsnArgLeuLeuSerProArgProSerLeuLeuThrPro  947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302  GAGGCCGAGGGCGGCGACAAGAACCGGCTGCTGTCCCCAAGGCCCAGCCTCCTCACCCCG  361

Qy   948  ThrGlyAspProArgAlaAsnAlaSerProGlnLysProLeuLeuAspLeuLysGlnLeuLys  967
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362  ACTGGGGACCCCCCGGGCCAATGCCTCACCCCAGAAGCCACTGGACCTGAAGCAGCTGAAG  421

Qy   968  GlnArg-AlaAlaAlaIleProProIleGlnValThrLysValHisGluProProArgGl  987
          |||||||     |||||||||||||||||||  |||||||||||||||||||||||||||
Db   422  CAGCGNAGCGGCTGCCATCCCCCCCCATC--GTCACCAAAGTCCATGAGCCCCCCCGGGA  478

Qy   987  uAspAlaAlaProThrLysProProAlaProProAlaProProProGlnAsnLeuGlnPr  1007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479  GGACGCAGCTCCCACCAAGCCAGCTCCCCCAGCCCCACCGCCAAAACCTGCAGCC  538

Qy  1007  oGluSerAspAlaProGlnGlnProGlySerSerProArgGlyLysSerArgSerProAl  1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539  GGAGAGCGACGCGCCCCTCAGCAGCCTGGCAGCAGCCCCCGGGCAAGAGCAGGAGCCCCGGC  598

Qy  1027  -----------------------AlaPheAlaAlaAlaGluAlaGl  1040
                                  |||||||||||||||||
Db   599  ACCCCCGCGCCGACAAGGAGGAGAAGAGCCTTGTGTTCTTCGGCAGCCGAGGCCCA  658

Qy  1040  nLysLeuProGlyAspProCysTrpThrSerGlyLeuProPheProValProProAr  1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   659  GAAGCTGCCTGGGGACCCCTGCTGGACTTCCGGGACTTCCGCCTCCCCGTGCCCCCCG  718

Qy  1060  gGluValIleLysAla-SerProHisAla-ProAspProSerAla-PheSerTyrAlaPr  1079
          |||||||||||||||  ||||||||    |||||||||||||||  |||||||||||||
Db   719  TGAGGTGATCAAGGCCCTCCATGCCCCCGACCCCTCAGCCCTTCTCCTACGGTCC  778

Qy  1079  oProGlyHisProLeuProLeuHisAspThr---AlaArgProValLeuProProAr  1098
          |||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db   779  ACCTGGTCACCCACTGCCCCTGCGGCCTCCATGACCCCTGCCCCCGTCCTGGCGCG  838

Qy  1098  gPro-ProThrIleSerAsnProProLeuIle-SerSerAlaLysHisPro  1115
          ||||    |||||||||||||||||||||||  |||
Db   839  CCCACCACCAACAATCTCCAACCGGCTCCCCTCATTCTCTCCTGCCAAGCACCC  892
```

RESULT 27
BM910704

LOCUS   BM910704   1038 bp   mRNA   linear   EST 12-MAR-2002
DEFINITION AGENCOURT 6617193 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454241
            5', mRNA sequence.
ACCESSION   BM910704

FEATURES   Location/Qualifiers
     source   1..1038
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454241"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 1038
Score: 1.95e-46   Matches: 277
Percent Similarity: 1346.50   Conservative: 11
Best Local Similarity: 82.76%   Mismatches: 36
Query Match: 79.60%   Indels: 24
DB: 10.19%   Gaps: 6
                    12

US-09-522-753-5 (1-2517) x BM910704 (1-1038)

```
Qy  2151  HisProGlnLeuSerAlaProAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer  2170
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11  CACCCCACAGCCAGTCAGCGGCACCCTGCCCGCCCCCCTGTACTCCTTCCCTGGGGCCCAGC   70

Qy  2171  CysProValLeuAspLeuArgArgProSerLeuTyrLeuProProSerLeuProProAspHis  2190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71  TGCCCCGTCCTGGACCTCCGGCGCCCCTCCCTGTACCTCCCCCCGCCCCCCAGTGACCCAT  130

Qy  2191  GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsn  2210
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131  GGTGCCCCGGCCCGTGGCTCCCCCCACACGCGAAGGGGGCAAGAGGTCTCCAGAGCCCAAAC  190

Qy  2211  LysThrSerValLeuGlyGlyGlnValGluAspGlyIleGluProValSerProGluGly  2230
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191  AAGACTTCGGTCTTGGGTGGTGGTCAGGTGGAGGACGGGGTATTGAACCTGTGTCCCCACCGG  250

Qy  2231  MetThrGluProSerArgMetGlySerProAlaValTyrProLeuLeuTyrArgAspGlyGlu  2250
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251  ATGACGGAGCCAGGGCACTCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGAA  310

Qy  2251  GlnThrGluProSerArgMetGlyAsnThrSerGlnProPro Ala  2270
          |||||||||||||||||||||||||||||
```

```
                    found through the I.M.A.G.E. Consortium/LLNL at:
                    http://image.llnl.gov
                    Plate: NDAM0055   row: j   column: 05
                    High quality sequence stop: 649.
FEATURES            Location/Qualifiers
      source        1..887
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:30143932"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_164"
                    /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
                    Non-normalized full-length enriched library from pooled
                    mouse embryonic limb, maxilla and mandible, day 10.5 and
                    11.5 (size selected for the 0.5-1 kb fragments) Cloned
                    directionally, priming method: Oligo-dT.   cDNA enrichment:
                    >1k bp, Average insert size 1.8k bp. Priming sequence:
                    5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T)3'. Tissue contributed
                    by, David Rowe. Library constructed by ResGen, Invitrogen
                    Corp."

ORIGIN
Alignment Scores:
Pred. No.:                                        887
Score:                      3.69e-46            1338.00
Percent Similarity:         89.00%              262
Best Local Similarity:      84.79%              13
Query Match:                10.12%              16
DB:                         14                 18
                                                5

US-09-522-753-5 (1-2517) x CA979881 (1-887)

Qy   1614  LeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAsp  1633
           ||||||||||||||||||||| :::  |||||||||||||||||||||||||||||||
Db      1  TTGCTCCGGGGCTGACTGGTGTGGACCTGTACCGGTCACATCCCATTGGCCTTTGAC       60

Qy   1634  ProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuPro  1652
           |||||||||||||||||||||||||||||:::
Db     61  CCCACCTCCATACCCGAGGGATCCCTCTGGAAGCAGCAGCCGCAGCCTACCTGCCC      120

Qy   1653  ArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProTyrLeuIleArgGly     1672
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CGGCACTTGGCCCCCAACCCCAGCCCCACACCTTACCCATCTTACCTACTCCGCGGC     180

Qy   1673  TyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSer  1692
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TACCCTGACACGGCGGCCCTGGAGAACCGCCAGACCATCATCAATGACTACATCACCTCG  240

Qy   1693  GlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGly  1712
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CAGCAGATGCACCACAACGTCGCCTGCCTCCGCATGGCCCAGCGTGCTGACATGCTGAGGGGT  300

Qy   1713  LeuSerProGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyTyrProArgGlyIleIle  1732
           |||||||||||||||||||||||||||||:::
Db    301  CTGTCACCCGAGAGTCCTCGCTGGCCTCAATTATGCGGTCTGGCCCAAGAGCATTATC  360

Qy   1733  AspLeuSerGlnValProHisLeuProValLeuProHisProGlyThrProThrProAla  1752
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GACCTGTCCCAAGTGCCACACCTGCCCGTGCTGCCGCCACCAACGCCAGGCACCCCCTGCC  420

Qy   1753  ThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPheSerSerArgHis  1772
           |||||||:::||||||||||||||||||||||||||||||||||||||||||
Db    421  ACCGGCCATCGACCCGGCCTTGCCTACCTCCCCACTGCCCCCCACCCTTCAGCAGCCGCCAC  480

Qy   1773  SerSerSerProLeuSerProGlyGlyProThrHisLeuThrHisProThrThrSer     1792
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  AGTAGCTCACCCGTCTCCCCAGGAGGCCCCACTCACCTAGTCAAACCACTGCCACATCT  540

Qy   1793  SerSerGluArgGluArgGluArgAspArgAspArgAspArgAspArgAspArgGluGluLys  1812
           ||||||||||||||||||||||:::|||||||||||||||||||||||
Db    541  TCATCGGAGCGGGAACGGGAGCAGGGAGACGTGAGCGGGAACGAGAC-----------------AAG  582
```

```
Db    311  CAGACGGAGCCCAGCAGGATGGGCTCCAGCAACACCACCGCCAGCGCCAGCC  370
Qy   2271  PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIle  2290
           |||||||||||||||||||||||||||||||||||||||||||||
Db    371  TTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAACCAAGAGATC  430
Qy   2291  AsnLysLysLeuAsnThrHisAsnArgAsnArgGluProGluTyrAsnIleSerGlnProGly  2310
           ||||||||||:|||:|||||||||||||||||||||||||||||||||||||||
Db    431  AACAAGAAGCTGAACACCCACACCCGAAATAGAGCCTGAATACAATATCAGCCAGCTGGG  490
Qy   2311  ThrGluPheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln  2330
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    491  ACGGAGGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCAG  550
Qy   2331  AlaValAlaGlnGluHisAlaSerThrArgMetGlyLeuGluAlaIleIleGluArgLysAlaLeu  2350
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    551  GCGGTGCAGGAACATGCCCGCACCAGATGGGCTGGAGGCCATAATTAGAAGGCACTC  610
Qy   2351  MetGlyLysTyrTyrAspTyrAspTrpGluGluSerProProLeuSerArgAlaAsnAlaPheAsnPro  2370
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    611  ATGGGTAAATATGACCAGTGGGAAGAGTCCCCGCGCTCAGCGCGCCAATGCTTTTAACCCT  670
Qy   2371  LeuAsnAlaSerAla-SerLeuProAlaAlaMetProIleThrAlaAla-AspGlyArgArgS  2390
           |||||||||||:::||||||||||:::   |||  |||||||| ||||||||
Db    671  CTGAATGCCAGTGCCGCCGCCCTGCCCCCTGCTATGCCCATAACCGTTGCTGACGGACGGA  730
Qy   2390  erAspHisThrLeuThrSerProGlyGly--GlyGlyLysAlaLysValSerGlyArg--  2408
           ||| ||| :||:||: ||||||||||||||||||||||||||:::
Db    731  GTGACCACACACTCACCTCCCCAGGTGGGCGGCGGGAAAGCCCAGGGTCTCTGGGCAAAC  790
Qy   2409  -ProSerSerArgGlysAlaLysSerProAlaProGlyLeuAlaSer---GlyAspArgPr  2427
           ||| ||:  :::
Db    791  CCCACAANCCGAAAGCCAATTCCCCGGCCCCCGGCCGGAATCTGGGGACCCGCC  850
Qy   2427  oProSerValHisSerGluGlyValHisSerGluGlyAspCys-------------ArgArgT  2443
           |||:  |||  :::|||   |||
Db    851  CCCCCTTTGGTTCTCCAGGGCACCTCCCGAGGGGCCACCTCTAAACCCCTGGGAAC  910
Qy   2443  hrProLeu--ThrAsnArgValTrp------GluAspArgProSerSerAlaGlySerT  2460
           ||| |||     ||:  |||
Db    911  GCCCTCTTCCCCAAACCTCTTTGTGGGGGCAAGGAACCCGCCCTTTNTCCCGGCGGCC  970
Qy   2460  hrProPheProTyrAsnProLeuIleMetAlaArgLeuGlnAlaGlyValMetAlaSerProP  2480
           |||:  ||||
Db    971  TTCCCCGCCCCATTTCCCCCCTA---------------------ACACCCC  1000
Qy   2480  roProProGlyLeu  2484
           |||
Db   1001  CCCCCCGGTCTCTC  1014
```

```
RESULT 28
CA979881
LOCUS       CA979881              887 bp    mRNA    linear    EST 06-JAN-2003
DEFINITION  AGENCOURT 11276590 NIH MGC 164 Mus musculus cDNA clone
            IMAGE:30143932 5', mRNA sequence.
ACCESSION   CA979881
VERSION     CA979881.1    GI:27512535
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 887)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
```

```
Qy   1813 SerIleLeuThrSerThrThrValGluHisAlaProIleTrpArgProGlyThrGlu 1832      Qy   877 GlyLysAspAlaGlyAlaAlaAlaGluAlaThrAlaGluThrAlaGluLysAlaGluLys 896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||                      |||:::  |||||    |:::  ||||||||||||||||||||||||| :::
Db   583 -TCCATCCTCACGTCTACCACTACAGTGGAGCATGTACCCATCTGAGACCTGGTACGGAG 642      Db     6 GGCAGGCAGGCGACCGNAGCCTGAGCCACCGGCCGGCGCCTCAG-GCAGAGAGAG-  63

Qy   1833 GlnSerSerGlySerSerGlySerGlyGlyGlyGlyGlySerGlySerSerArgProAla 1852      Qy   897 GluGlyGlySerGlyArgGlyAlaThrArgAlaLysSerSerGlyAlaProGlnAspSerAsp 916
          |||:::|||||||:::                              :::             ||           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   643 CAGAGCAGCGGGGCT-----------------------------GGGGGCAGCAGCCGCCCCGCC 678      Db    64 GAGGGCGGGAGCGGCAGGGCGCCAGGAGTCGGGCGCCCCAGACAGCGAC 123

Qy   1853 SerHisSerHisAlaHisGlnHisSerProIleSerProArgThrArgGlnAspAlaLeuGln 1872      Qy   917 SerSerAlaThrCysSerAlaAspGluValAspGluLysAspGlyLysAsnArg 936
          |||:::|||      |||||||||||:::                                                   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   679 TCCCACACC-------CACCAGCACTCGCCCATCTCCCCGGACAGCCCAGAGACGCCTTGCAG 732      Db   124 TCCAGTGCTACCTGCAGTGCAGAGGTGCAGAGGTGATGAGGCCGGCGACAAGAACCGG 183

Qy   1873 GlnArgProSerValLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluPro 1892      Qy   937 LeuLeuSerProArgProSerLeuThrProThrGlyAspProArgAlaAsnAlaSer 956
          ||||||||||||||||||||||||||||||||||||||:::  :::|||||||||||||                  |||||||||||||||||||||||||||||||||||||||||||||||
Db   733 CAGAGGGCCCAGTGTGTGCTGCACAACACGGGCATGAAGGGCGTGGTCACCGCCGTGGAACCC 792      Db   184 CTGCTGTCCCCAAGGCCCAGCCTCCTCACCCCGACTGGCGACCCCCGGGCCAATGCCTCA 243

Qy   1893 SerIleSerProThrValLeuArgSerThrSerThrSerSerProValArgProAlaAlaThr 1912      Qy   957 ProGlnLysAspLeuLeuLysGlnArgAlaLeuLysAlaAlaIleAlaIleProProIle 976
          |||||||||||||||||||||||||||||:::|||||:::|||||                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   793 GGCACGCCCACGGTCCTGAGGTCCCCACTCACCTCTTCGCCTGTTCCGCCCCAGCTGCCACA 852      Db   244 CCCCAGAAGGCCACTGACCTGAGCAGCAGCGAGCAGCGGCTGCATCCCCCCCATC 303

Qy   1913 PheProPro--AlaThrHisCysProPro 1920      Qy   977 GlnValThrLysValHisGluProProArgGluAspAlaAlaAlaProThrLysProAlaPro 996
          ||||||      :::||||||                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   853 TTTCCAACCTGGCCACCCAACTGCCCC 879      Db   304 CAGGTCACCAAAGTCCATGAGCCCCCGGGGAGGACGCAGCTCCACCAAGCCAGCTCCC 363
```

```
RESULT 29
BX390462                                                                                 Qy   997 ProAlaProProProGlnAsnLeuGlnProGlnSerAspAlaProGlnAsnProGly 1016
LOCUS         BX390462                926 bp    mRNA    linear    EST 23-JUN-2003                ||||||||||||||||||||||||||||||||||||||||||||||||||||
DEFINITION    BX390462 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA        Db   364 CCAGCCCCACCGCCCACCGCCAAACCTCAGCCCGGAGAGGACGGGCCCTCAGCAGCCTGGC 423
              clone CS0DI056YJ24 5-PRIME, mRNA sequence.
ACCESSION     BX390462                                                                  Qy  1017 SerSerProArgGlyLysSerArgSerProAlaProProAlaAspLysSerGlu------- 1033
VERSION       BX390462.1   GI:32165291                                                          ||||||||||||||||||||||||||||||||||||||||||||||
KEYWORDS      EST.                                                                      Db   424 AGCAGCCCCGGGGCAAGAGCAGGAGCCCGACCCCCGCCGACAAGAGGAGAAGCCT 483
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens                                                              Qy  1034 --------AlaPheAlaAlaGluAlaGlnLysLeuProGlyValAspProProCysTrp 1049
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;                   ||||||||||||||||||||||||||||||||||||||||
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.                 Db   484 GTGTTCTTCGCAGCCTTCGAGCCCAGAAGCTCGCTGGGGGACCCCCCTTGCTGG 543
REFERENCE     1  (bases 1 to 926)
AUTHORS       Li,W.B., Gruber,C., Jessee,J. and Polayes,D.                              Qy  1050 ThrSerGlyLeuProPheProValProProArgGluValIleLeuLysAlaSerProHisAla 1069
TITLE         Full-length cDNA libraries and normalization                                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
JOURNAL       Unpublished (2001)                                                        Db   544 ACTTCCGGCTGCCTCCCTGCCCCGTGCCCCCCCGAGGTGATCAAGGCCTCCCCGCATGCC 603
COMMENT       Contact: Genoscope
              Genoscope - Centre National de Sequencage                                 Qy  1070 ProAspProSerAlaPheSerTyrAlaProProGlyHisProLeuProLeuGlyLeuHis 1089
              BP 191 91006 EVRY cedex - France                                                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
              Email: secref@genoscope.cns.fr, Web : www.genoscope.cns.fr                Db   604 CCGGACCCCTCAGCCTTCTCCTACGCTCCACCTGGTCACCCCACTGCCCCTGGGCCTNCAT 663
              Library was constructed by Life Technologies, a division of
              Invitrogen. This sequence belongs to sequence cluster 7437.f,             Qy  1090 AspThrAlaArgProValLeuProArgProProThrIleSerArgAsnProProLeuLeuIl 1109
              Contact: Feng Liang Email : fliang@lifetech.com URL :                             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
              http://fulllength.invitrogen.com/ InvitroGen Corporation 1600            Db   664 GACACTGCCCGGCCCCGTNCTGCCCGCGCCCACCCACCATNTNCAACCCGNCTCNCCTNAT 723
Faraday Avenue Genoscope sequence ID : CS0BAG032ZG08_CS03039_1'.
FEATURES      Location/Qualifiers                                                       Qy  1109 eSerSerAlaLysHisProSerValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMe 1129
source        1..926                                                                            |||||||||||||||||||||||||||||||||||||||||||||||||||
              /organism="Homo sapiens"                                                  Db   724 CTNCCTGCCAAGCACCCCAGCGTCCTGGAGAGGCAAAATAGGTGCCATCNCACCACAGGAAG 783
              /mol_type="mRNA"
              /db_xref="taxon:9606"                                                     Qy  1129 tSerValGlnLeuHisValProTyrSerGluHisAlaLysAlaProValGlyProValTh 1149
              /clone="CS0DI056YJ24"                                                             ||||||||||||||||||||||||||||||||||||||||||||||||
              /tissue_type="PLACENTA COT 25-NORMALIZED"                                 Db   784 -TNNGTNCAGCTNCACGTCCCGTACTCAGAGCATCCCAAGCC--CCCGTGGGCCCCTGTCAC 841
              /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
              /note="1st strand cDNA was primed with a NotI-oligo(dT)                   Qy  1149 rMetGlyLeuProLeuProMetAspProLysLeuAlaProPheSerLysValLysGl 1169
              primer. Five prime end enriched, double-strand cDNA was                           ||||||||||||| :::||||||||||||||||||||||||||||||||||||
              digested with Not I and EcoR V"                                            Db   842 CATGGGGCTGCCCCTGCCCATGGACCCCAAAAAGTGGCACCCTTCAGCCGGAGTGAAGAA 901
              sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN                                                                                   Qy  1169 nGluGlnLeuSerProArgGlyGln 1177
                                                                                                 |||||||  |||:::
                                                                                         Db   902 -GAACAGCTGTCCCCAGGGGGCAGG 925
```

```
Alignment Scores:                                                                        RESULT 30
Pred. No.:                                    Length:          926                        BQ711119
Score:                 5.66e-46               Matches:         276                        LOCUS         BQ711119
Percent Similarity:    1334.00                Conservative:    4                          DEFINITION    BQ711119 AGENCOURT_7977236 NIH_MGC_113 Homo sapiens cDNA
Best Local Similarity: 90.61%                 Mismatches:      21                                       clone IMAGE:6215773
Query Match:           89.32%                 Indels:          13                                       5', mRNA sequence.
DB:                    10.09%                  Gaps:            1                          ACCESSION     BQ711119
                       13                                                                 VERSION       BQ711119.1   GI:2185018
                                                                                          KEYWORDS      EST.
US-09-522-753-5 (1-2517) x BX390462 (1-926)                                               SOURCE        Homo sapiens (human)
```

BQ711119　　　　875 bp　　　mRNA　　linear　　EST 16-JUL-2002

```
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 875)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Dr. Mark Watson
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM2385  row: d  column: 14
          High quality sequence stop: 668.
FEATURES             Location/Qualifiers
     source          1..875
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6215773"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G).Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
```

ORIGIN

```
Alignment Scores:
Pred. No.:                                  Length:          875
Score:              6.44e-46                Matches:         252
Percent Similarity: 1332.00                 Conservative:    0
Best Local Similarity: 100.00%              Mismatches:      0
Query Match:        100.00%                 Indels:          0
DB:                 10.08%                  Gaps:            0
                    13
```

US-09-522-753-5 (1-2517) x BQ711119 (1-875)

```
Qy  2266 SerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSer 2285
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3 AGCCAGCCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCC 62

Qy  2286 LysLysGlnIleLysAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsn 2305
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    63 AAGAAGCAAGAGATCAACAGAGAGTCAACACCCACAACCGGAATGAGCCTGAATACAAT 122

Qy  2306 IleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMet 2325
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 ATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATG 182

Qy  2326 ThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyMetGluAlaIle 2345
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 ACCTATAGAAGCCAGGCGGTGCAGGAACATGCCAGCACCAACATGGGCATGGAGGCCATA 242

Qy  2346 IleArgLysAlaLeuMetGlyTyrGlyLeuTyrAspGlnTrpGluSerProProSerAla 2365
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 ATTAGAAAGGCACTCATGGGTAAATATGACCAGTGGGAAGAGTCCCCGCCCAGCGCC 302

Qy  2366 AsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAla 2385
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 AATGCTTTTAACCCTCAATGCCAGTGCCAGCTGCCAGCTGCCATGCCCATAACCGCT 362

Qy  2386 AlaAspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysVal 2405
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 GCTGACGGACGGAGTGACCACACACTCACCTCACCCGGAGGTGGCGGGGAAGGCCAAGGTC 422
```

(Continued right column)

```
Qy  2406 SerGlyArgProSerSerArgLysAlaLysSerProAlaProGlyLeuAlaProGlyAsp 2425
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423 TCTGGCCAGACCCAGCCAGCCGAAAAGCCAAGTCCCGGCCCCGGGCCTGGCATCTGGGAC 482

Qy  2426 ArgProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeu 2445
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483 CGGCCACCCTGTCTCCTCAGTGCACTGCACTGGAGGGAGACTGCAACCGCCGGACGCCGCTC 542

Qy  2446 ThrAsnArgValTrpGluAspArgProSerArgAlaGlySerThrProPhePheTyrAsn 2465
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543 ACCAACCGGGTGTGGGAGGACAGGCCCTCGTCCCAGGTTCCACGGTTCCCTACAAC 602

Qy  2466 ProLeuIleMetArgLeuGluIleAlaGlyValMetAlaSerProProProGlyLeuPro 2485
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   603 CCCCTGATCATCGGCTGCAGGCGGGTGTCATGGCTTCCCCACCCCCACCGGGCCTCCCC 662

Qy  2486 AlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpAspGluProLysProLysLeu 2505
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   663 GCGGGCAGCGGGCCCCTGGCTGGCCCACCACGCTGGGACGAGGAGCCCAAGCCACTG 722

Qy  2506 LeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
              |||||||||||||||||||||||||||||||||||
Db   723 CTCTGCTCGCAGTACGAGACACTCTCCGACAGCGAG 758
```

RESULT 31

```
BU224569
LOCUS       BU224569    842 bp    mRNA    linear    EST 26-NOV-2002
DEFINITION  603400211F1 CSEQCHN23 Gallus gallus cDNA clone CHEST293f3 5', mRNA
            sequence.
ACCESSION   BU224569
VERSION     BU224569.1  GI:25459595
KEYWORDS    EST.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 842)
AUTHORS     Boardman,P.E., Sanz-Ezguerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE       A Comprehensive Collection of Chicken cDNAs
JOURNAL     Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE     22335534
PUBMED      12445392
COMMENT     Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology
            (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES             Location/Qualifiers
     source          1..842
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="CHEST293f3"
                     /dev_stage="22"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN23"
                     /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
```

```
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
```

```
Db        808 --------ATGCCCCG-AGCAGCCAAGAGAGAGAAAGATGAACAA 842
              ||||||  |||||||||||||||||||||||||||||||||
```

RESULT 32
```
BM560912
LOCUS        BM560912                1094 bp    mRNA    linear   EST 20-FEB-2002
DEFINITION   AGENCOURT_6566208 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5550790
             5', mRNA sequence.
ACCESSION    BM560912
VERSION      BM560912.1  GI:18805729
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1094)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM2263  row: h  column: 23
             High quality sequence start: 30
             High quality sequence stop: 676.
FEATURES             Location/Qualifiers
     source          1..1094
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5550790"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6;  Site_1: NotI;
                     Site_2: SalI; Cloned unidirectional.  Primer: Oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."
ORIGIN
```

Alignment Scores:
```
Pred. No.:                                 1094
Score:                     1.46e-45        Length:               1094
Percent Similarity:        1326.00         Matches:               284
Best Local Similarity:     81.99%          Conservative:           12
Query Match:               78.67%          Mismatches:             34
DB:                        10.03%          Indels:                 33
                           12              Gaps:                    5
```

US-09-522-753-5 (1-2517) x BM560912 (1-1094)

```
Qy   613 SerArgTrpThrGluGluGluMetGluThrAlaLysLysLysLeuGluHisGlyArg 632
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41 TCTCGCTGACAGAAGAAGAAATGGAAACAGCCAAGAAGAAAGTCTCCTGGAACACGGCCGGC 100

Qy   633 AsnTrpSerAlaIleAlaArgMetValGlySerLysThrValSerGlnCysLysAsnPhe 652
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 AACTGGTCGGCCATCGCCCGGATGGTGGGCTCCAAGACTGTGTCGCAGTGTAAGAACTTC 160

Qy   653 TyrPheAsnTyrLysLysArgGlnAsnLeuAsnLeuLeuGluHisGlnHisLysLeuLys 672
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 TACTTCAACTACAAGAAGAGGCAGAACCTCGATGAGATCTTGCAGCAGCACAAGCTGAAG 220

Qy   673 MetGluLysGluArgAsnLeuArgArgArgGlyLysLysLysAlaAlaProAlaAlaSerGlu 692
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 ATGGAGAAGGAGAGGAACCGCCGGAGGAAGAAGAAGGCCGCCCGGCCGCCAGCGAG 280

Qy   693 GluAlaAlaPheProValValGluAspGluLysMetGluAlaAlaSerGlyValSerGly 712
```

US-09-522-753-5 (1-2517) x BU224569 (1-842)

Alignment Scores:
```
Pred. No.:
Score:                     6.47e-46        Length:               842
Percent Similarity:        1331.50         Matches:              269
Best Local Similarity:     93.24%          Conservative:           7
Query Match:               90.88%          Mismatches:            18
DB:                        10.08%          Indels:                 2
                           13              Gaps:
```

```
Qy   231 TyrAspGluAsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyPro 250
         |||||||||||||||||||:||||||||||||||||||||||||||
Db     2 TACGACGAGAACAGGAAGAAGGCAGAAGCTGCCCACCGCATCCTGGAAGGACTGGGTCCA 61

Qy   251 GlnValGluLeuProLeuTyrArgGlnProSerAspThrArgGlnTyrHisGluAsnIle 270
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 CAGGTGGAGCTGCCGTTATACACCAGCCTTCAGACACCAGGCAATACCATGAAAAACATT 121

Qy   271 LysIleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAla 290
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 AAAATAAACCAGGCAATGCGGAAGAAGTTAATTTTGTACTTTAAGAGGAGAAATCATGCT 181

Qy   291 ArgLysGlnTrpLysGlnLysPheCysByrLeuArgTyrAspGlnLeuMetGluAlaLeuGlu 310
         |||||||||||:||||||||||||||||||||||||||||||||||||||
Db   182 CGCAAGCAGTGGGAACAGAAGTTCTGCCAGCGCTATGACCAGCTGATGGAGGCCTGGGAA 241

Qy   311 LysLysValGluArgIleLeuGluAsnAsnProArgArgArgAlaLysGluSerLysValArg 330
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 AAGAAAGTGGAGCGCATTGAAAATAACCCCCGGCCGTCCAAGGAGAGCAAAGTGCGG 301

Qy   331 GluTyrTyrGlyLysGlnPheGluLeuArgLysGluArgGluLeuGlnArgArgMet 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 GAATACTATGAGAAGCAATTCCCTGAGATCCGGAAGCAGCGGGAACTGCAAGAGCGCATG 361

Qy   351 GlnSerArgValGlyGlnArgGlySer---GlyLeuSerMetSerAlaAlaArgSerGlu 369
         |||||:|::|||||||::|||:     |||||||||||||||||||:||
Db   362 CAGAGCAGGGTAGGACAGAGGGGCAGCGGGGGCCTCTCCATGTCGGCTGCCGCCAGTGAG 421

Qy   370 HisGluValSerGluIleIleGluAspGlyLeuSerGluGlnGluLeuAsnLeuGluLysGlnMet 389
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 CACGAGGTGTCGGAGATCATCGATGGGCTTCAGAGACAAGAAGAACCTGGAGAAGCAGATG 481

Qy   390 ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnAsnArgIleLeuLysPhe 409
         ||||||||||:|||||||||||||:||||||||||||||||||||||||:||||
Db   482 CGCCAGCTCGCCGTCATCCCACCCATGCTCTATGATGCTGACCAGCAGCGCATTAAAATTC 541

Qy   410 IleGluAsnMetAsnGlyLeuMetAlaAspProMetLysMetLysValTyrLysAspArgGlnValMet 429
         ||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db   542 ATCAACATGAACGGCCTCATGGCAGACCCCATGAAGTTCTACAAGGACCGGCAGTGATG 601

Qy   430 AsnMetTrpSerGluGluGlnLysThrPheGluGluArgGluMetGlnHisProLys 449
         ||||||||:|||||||||||||||||||||||||||||||||||||||
Db   602 AACATGTGGAGCGAGCAGCAGAAGACAGAGAGCTTCGAGGAAAGAAGTTCATGCAGCACCCAAAG 661

Qy   450 AsnPheGlyLeuIle-AlaSerPheLeuGluArgLysThrVal-AlaGluCysValLeuT 469
         |||||||||:||: :||:|::|||||||||||||||||:||:|||||||||||||
Db   662 AACTTTGGCCTGATAGGTGCCTCCTTTCTGGAGAAGAAGACGGTTGCGACTGTGCTTCCTGT 721

Qy   469 TyrTyrPheLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgSerTyrArgA 489
         |||||:|||||||||||||||||||||:||::|||||||||
Db   722 ACTACTACCTGACCAAGAAGAATGAGAACTACAAGAACCTCGTGAGGAGAATTACCCGGC 781

Qy   489 rgArgGlyLysSerGlyLysGluAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 509
         ||||||||||:||||||||||||||||||||
Db   782 GGCGTGGGAAAAGCCAGCAGCAGCAA--------------------- 807

Qy   509 lnGlnProMetProArgSerSerGlnGluGluLysAspGluLys 523
```

```
Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13559  row: g  column: 10
High quality sequence stop: 722.

FEATURES       Location/Qualifiers
  source       1..905
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:6178977"
               /sex="male"
               /tissue_type="dorsal root ganglia"
               /dev_stage="adult, 36 yr"
               /lab_host="DH10B"
               /clone_lib="Lupski dorsal root ganglion"
               /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
               NotI; Site 2: SalI; cDNA made by oligo-dT priming.
               Directionally cloned using the following adaptors:
               5'-TCGACCCACGCGTCCG-3' and
               5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected >
               1 kb for average insert length 1.7 kb. This is a primary
               library, non-amplified. Library constructed by Life
               Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
               College of Medicine) and is available through Life
               Technologies."

ORIGIN

Alignment Scores:
Pred. No.:              1.24e-45        Length:          905
Score:                  1325.50         Matches:         255
Percent Similarity:     77.27%          Conservative:    0
Best Local Similarity:  77.27%          Mismatches:      0
Query Match:            10.03%          Indels:          75
DB:                     13              Gaps:            1

US-09-522-753-5 (1-2517) x BQ897825 (1-905)

Qy  2188 ProAspHisGlyAlaAlaProAlaArgGlySerProHisGlyGlyLysGlyGlyLysArgSerPro 2207
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CCGGACCCATGGTCCCCGGCCCGTGGCTTCCCCCCACACGCGAACGGGGCAAGAGGTCTCCA  60

Qy  2208 GluProAsnLysThrSerValLeuGlyLysGlyGlyIleGluAspProValSerPro 2227
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GAGCCCAAACAAGACGTCGGTCTTGGGTGTGGTGGGATTGGAGACGGTATTGAACCTGTGTCCCCA 120

Qy  2228 ProGluGlyMetThrGlyProGlyHisSerArgSerArgAlaValTyrProLeuLeuTyrArg 2247
         ||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCGGAGGGCCATGACGGAGCCAGGGCACTCCCGGAGTGCGTGTGTGGCTGTGCTGTACCGG 180

Qy  2248 AspGlyGluGlnThrGlyProSerArgMetGlySerLysSerProGlyAsnThrSerGln 2267
         |||||||||||||||||||||||||||||||||||||||||||
Db   181 GATGGGGAACAGACAGGACGGAGCCAGGAGGATGGGCTTCCAAGTCTCCAAGCAACACCAGCCAG 240

Qy  2268 ProProAlaPhePheSerLeuThrGluSerAsnSerAlaMetValLysSerLysPro 2287
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCAAGAAG 300

Qy  2288 GlnLysAsnLysLysLysAsnThrHisAsnArgAsnGluProTyrAsnIleSer 2307
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAAGAGATCAACAAGAAGTCAACACCCCACACCGAATGAGCCTGAATATCAGC 360

Qy  2308 GlnProGlyThrGlnIlePheAsnMetProAlaIleThrGlyThrMetThrTyr 2327
         |||||||||||||||||||||||||||||||||||||||||
Db   361 CAGCTGGGACGGACGGAGATCTTCAATATGCCCCATCACCGGAACA------------ 405

Qy  2328 ArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArg
Db   405 -------------------------------------------------------------
```

```
Db   281 GAGGCTGCATTCCCGCCGTGGTGGAGGATGGAGGAGATGAGGCGTTCGGGCGTGAGCGGGA 340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   713 AsnGluGluValGlyMetValGlyGluAlaGluLeuAlaLeuHisAlaSerGlyAsnGluValPro 732
         |||||||||||||||||||||||||
Db   341 AATGAGGAGGAGAGATGGTGGAGGAGGCTGAA----------------------------- 370

Qy   733 ArgGlyGlyCysSerGlyProAlaThrValAsnAsnSerSerAspThrGluSerIlePro 752
         ----------GCCACTGTCAACACAGCTCAGACACACCGAGCACATCCCC 409
Db   371

Qy   753 SerProHisThrGluAlaLysAspThrGlyGlnAsnGlyProLysProProAlaThr 772
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   410 TCTCCTCCACACTGAGGCCGCCAAGGACACACAGGGCAGAATGGGCCAGCCCCCAGCCACC 469

Qy   773 LeuGlyAlaAspGlyLeuProProGlyProProThrProPro-ArgArgThrSerArgAl 792
         |||||||||||||||||||||||||||||||||||||||||||
Db   470 CTGGGGCCGCCGACGGGCCCACCCCCCACCACCGGAGGACATCCCGGCC 529

Qy   792 aProIleGluProThrProAlaSerGluAlaThrGlyAlaProThrProProProAlaPr 812
         ||||||||||||||||||||||||||||||||||||||
Db   530 --CCCACTGAGCCCACCCCCGGCCCTCTGAAGCCACCGGAGCCCCTTACCGCCACCAGGACC 588

Qy   812 oProSerProSerAlaProProProValValProLysGluGluGluGluThrAl 832
         ||||||||||||||||||||||||||||||||||||||||||
Db   589 CCCATCGCCCCTCTGCACCTCCTCGTCGTGTCCCCAAGGAGGAAGAAGAAGAGGAGACCGC 648

Qy   832 aAlaAlaProAlaProValGluGluGluGlyLeuGlyGlnLeuLysProProAlaAlaGluGluLeuAl 852
         ||||||||||||||||||||||||||||||||||||
Db   649 AGCAGCCGCCCCAGTGAGGAGGAGGGACGGACANAAACCCCCCCGGCTGAGGAGCTGGC 708

Qy   852 aValAspThrGlyLysAlaGluGluProValLysSerThrGluCysThrGlyGluAlaGluGl 872
         ||||||||||||||||||||||||||||||||||||||||
Db   709 AGTGGACACAGGAAAGGCCGAGGAGCCCGTCAAGAGCACCGAGTGCACCGGAGAGGCCCAGGA 768

Qy   872 uGlyProAlaLysGlyLysGlySerGlyTyrArgGlyAlaLysAlaGluGlyAlaGluLeuAl 892
         |||||||||||||||||||||||||||||||||||||||||||||||||       ::
Db   769 GGGGGCCGGCCAA--GGTAAGGACCCCGGCTCAGGAGGCCCGCTGAAGGGCCCGCTCA 827

Qy   892 sAlaGluLys---LysGlyLeuGlySerGlyTyrArgGlyAlaThrThrAla-LysSerSerGlyA 911
         ::::||||||       ||| ::::||||||||::|||||||||||||
Db   828 AGGCACACGGAACAGAGAGGGCCGCAGCGGCCAAGGGCCCGCAGAAGTTCGGGGGG 887

Qy   911 laProGlnAspSerAsp-SerSerAlaThrCysSerAlaAspGluValAspGlu-AlaGl 930
         ||||||||||||||||       :::::::::::::||||||
Db   888 CCCCCCGGCCCCAAGGAGCACCTCAGGGGGCCAGGGGCAACCAGGGGGAATTAGGCCGA 947

Qy   930 uGly---GlyAspLysAsnArgLeuLeuSerProArg-----------ProSe 943
         |||       :::              ||||||
Db   948 AGGGGCGGCCCAACACAACCCGGTTGTTGGTCCCACAGGGGCCCGGTCTTTCCTCCCCCCT 1007

Qy   943 rLeuLeuThrProThrGlyAspProArgAlaAlaAsnAlaSerProGlnLysProLeu 961
         ||||||       |||||
Db  1008 ATTGGCCCGACCCCGGTCCCCTAGCGT-------GCCCCCCCCCTAACGCCCCTG 1056
```

```
RESULT 33
BQ897825
LOCUS         BQ897825    905 bp    mRNA    linear    EST 16-AUG-2002
DEFINITION    AGENCOURT 8113194 Lupski dorsal root ganglion Homo sapiens cDNA
              clone IMAGE:6178977 5', mRNA sequence.
ACCESSION     BQ897825
VERSION       BQ897825.1  GI:22289839
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 905)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
```

```
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
       Site_2: EcoRI; cDNA made by oligo-dT priming.
       Directionally cloned into EcoRI/XhoI sites using the
       following 5' adaptor: GGCACGAG(G). Library constructed by
       Ling Hong in the laboratory of Gerald M. Rubin (University
       of California, Berkeley) using ZAP-cDNA synthesis kit
       (Stratagene) and Superscript II RT (Life Technologies).
       Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:              1.85e-45          Length:           882
Score:                  1321.00           Matches:          266
Percent Similarity:     94.74%            Conservative:     4
Best Local Similarity:  93.33%            Mismatches:       9
Query Match:            10.00%            Indels:           6
DB:                                       Gaps:             0

US-09-522-753-5 (1-2517) x BU180236 (1-882)

Qy   1334 AlaIleProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGly 1353
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCATCCCGCGGAGCGACACAGCCCCACCACTCAAAGAGCAGCACCACATCCGCGGG 60

Qy   1354 SerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArg 1373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCATCACACAAGGGATCCCTCGGTCCTACGTGGAGGCACAGGAGGACTACCTGCGTCGG 120

Qy   1374 GluAlaLysLeuLeuLysArgGluGlyTyrAspThrProProProSerArgAspLeu 1393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAGGCCAAGCTCCTAAAGCGGGAGGGCACCGCCCTCCGCCCCACCCGCTCACGGGACCTG 180

Qy   1394 ThrGluAlaTyrLysThrGlnAlaLeuGlyLysProLeuLysProAlaHisGluGly 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACCGAGGCCTACAAGACGCAGGCCCTGGGCCCCCTGAAGCTGAAGCCCATGAGGGC 240

Qy   1414 LeuValAlaThrValLysGluAlaGlyArgSerIleHisGluIleProArgGluLeu 1433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CTGGTGGCCACGGTGAAGGAGCCGGGCCGCTCAATCCATGAGATCCCGCGAGAGCTG 300

Qy   1434 ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluGlyIleThrGln 1453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGGCACACGCCCGAGCTGCCCCTGGCCCCGCGGCCGCTCAAGGAGGGCTTCATCACGCAG 360

Qy   1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlyLysLysHisAspVal 1473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACTGGCTCCAAAAAGCACGACGTA 420

Qy   1474 ArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAspValMet 1493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGCTCCCTCATCGGCAGCCCGGGCCGCACCTTCCCACCCGTCCACCCGCTGGATGTGATG 480

Qy   1494 AlaAspAlaArgAlaLeuGluArgAlaAlaCysTyrGluSerLeuLysSerArgProGly 1513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGAGCCTGAAGAGCCGGCCCAGGG 540

Qy   1514 ThrAlaSer-SerSerGlyGlySerIleAlaArgGlyAlaProValIleValProGluLeu 1533
          |||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACCGCCAGCAGGTTCGGGGGCTCCATTGCGCGGGGCGCCCCCGTCATTGTGCCTGAGCT 600

Qy   1533 uGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAla-ProPheAlaGlyH 1553
          |||||||||||||||||||||||||||||||||||||||||||   |||||||||||||
Db    601 GGGTAAGCCGCGGCAGAGCCCGCTGACCTATGAGGACCACGGCGGGCACCCTTTGCGGCC 660

Qy   1553 isLeuProArgGlySerProValThrMetArgGluProArgLeuGlnGluGlyLys 1573
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACCTCCCCAGGAGTTCGCCCGTGGACCCCGGGAGCCCACGCCCGCCTCAGGAGGCA 720

Qy   1573 erLeuSerSerLysAlaSer-GlnAspArgLysLeuThrSerThrProArgGluIle 1592
          ||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db    721 GCCTTTTCTCCCCAGGCTCCCCAGGACCGAAAGCTCACGTCAACGCCTCGGGAGATC 780

Qy   1593 AlaLysSerProHis-SerThrValProGluHisHisProHis-ProIleSer-ProTyr 1611
```

Qy   2348 LysAlaLeuMetGlyLysTyrAspGlnTrpGluSerProLeuSerProProLeuSerAlaAsnAla 2367
Db   405 --- 405

Qy   2368 PheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAsp 2387
Db   405 -- 405

Qy   2388 GlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysAlaLysValSerGly 2407
          ||||||||||||||||||||||||||
Db   406 -----------------------------GGTGCGGCGGAGGCCAAGGTCTCTGC 435

Qy   2408 ArgProSerSerArgLysSerAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgArgPro 2427
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 AGACCCAGCAGCCGAAAAGCCAAGTCCCCGGCCCCGGCCTGGCCATCTGGGACCGGCCA 495

Qy   2428 ProSerValSerSerValHisSerGluGlyValGlyAspCysAsnArgArgThrProLeuThrAsn 2447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 CCCTCCGTCTCCTCAGTGCACTCCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAAC 555

Qy   2448 ArgValTrpGluAspArgProSerSerArgAlaGlySerThrProPheProTyrAsnProLeu 2467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556 CGCGTGTGGGAGGACAGGCCCTCGTCCCGAGGTTCCACGCCATTCCCCTACAACCCCCTG 615

Qy   2468 IleMetArgLeuGlnLeuAlaGlyValMetAlaSerProProProGlyLeuProAlaGly 2487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   616 ATCATGCGGCTGCAGGCGGGTGTCATGGCTTCCCACCCCACCCGGGCCTCCCCGCGGGC 675

Qy   2488 SerGlyProLeuAlaGlyProHisHisAlaTrpAspGluProLysProLysProLeuLeuCys 2507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   676 AGCGGGCCCCTCGCTGGCCCCCACCACGGCCTGGGACGAGCCCAAGCCACTGCTCTGC 735

Qy   2508 SerGlnThrGlyGluThrGluSerAspSerGlu 2517
          |||||||||||||||||||||||||
Db   736 TCGCAGTACGAGACACTCTCCGACACGCAG 765

```
RESULT 34
BU180236
LOCUS       BU180236              882 bp   mRNA      linear     EST 04-SEP-2002
DEFINITION  AGENCOURT_8042921 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6084555
            5', mRNA sequence.
ACCESSION   BU180236
VERSION     BU180236.1  GI:22694220
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 882)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2316  row: i  column: 04
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1..882
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6084555"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
```

```
Db   781 GCCAAGGCCCGAACAACTACCGTGCCCGAGATCAACCCAACCCAATTTCGGCCCTAT 840

Qy  1612 Glu 1612
         |||
Db   841 GAA 843

RESULT 35
BX368972
LOCUS       BX368972           943 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  BX368972 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL010YH07 5-PRIME, mRNA sequence.
ACCESSION   BX368972
VERSION     BX368972.1  GI:30449776
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 943)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polyes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6282.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAE029ZG02_AE02704_2&cluster=6282.r.
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAE029ZG02_AE02704_2.
FEATURES             Location/Qualifiers
     source          1..943
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DL010YH07"
                     /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
                     25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with the Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:             4.49e-45        Length:        943
Score:                 1312.50         Matches:       273
Percent Similarity:    80.87%          Conservative:  6
Best Local Similarity: 79.13%          Mismatches:    32
Query Match:           9.93%           Indels:        36
DB:                    13              Gaps:          5

US-09-522-753-5 (1-2517) x BX368972 (1-943)

Qy  1579 AlaSerGlnAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSer 1598
         |||||||||:::||||||||||||||||||||||||||:::||||||||
Db     3 GCATCCCAGAACCG-AAGCTGACGTCCGACGCCTCGTGAGGACGCCAAGTCCCGCCACAGC 61

Qy  1599 ThrValProGluHisHisProIleSerProIleTyrGluHisLeuLeuArgGlyVal 1618
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 ACCGTGCCCGAGCACCACCACCCCACCCCATCTGCCCCATGAGCACCTGCTTCGGGCGGTG 121

Qy  1619 SerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIlePro 1638
         ||||||||||||||||||||||||||||||||||||||||||
Db   122 AGTGGCGTGGACCTGTATCGCAGCCACATCCCCCTGGCCTTCGACCCCACCTCCATACCC 181
```

```
Qy  1639 ArgGlyIleProLeuAspAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
         |||||||||||||||||||||||||||||||||::: |||||||||||
Db   182 CGCGGCATCCCTCTTGGACGCAGCCGCTGCCTACTACCTGCCCCACACCTGGCCCCCAAC 241

Qy  1659 ProThrTyrProHisLeuTyrProTyrLeuIleArgGlyTyrProAspThrAlaAla 1678
         ||||||||||||||||||||||||||||||||||||||||||
Db   242 CCCACCTACCCGCACCTGTACCCACCCTACCTCATCCGCGGCTACCCCGACACGGCGGCG 301

Qy  1679 LeuGluAsnArgGlnThrIleLeuAsnAspTyrIleThrSerGlnMetHisHisAsn 1698
         |||||||||||||||||||||||||||||||||||||
Db   302 CTGGAGAACCGACAGACCATCATCAATGACTACATCACCTCGCAGCAGATGCACCACAAC 361

Qy  1699 ThrAlaThrAlaMetAlaGlnAsnArgAlaAspMetLeuArgGlyLeuSerProArgGluSer 1718
         |||||||||||||||||||||||||||||||||||||||||||||
Db   362 GCGCGACCCACCGCCATGGCCCAGCGAGCTGATATGCTGAGGGGGCCTCTCGCCCGCGAGTCC 421

Qy  1719 SerLeuAlaLeuAsnTyrAlaAlaGlyGlyTyrIleIleAspProLeuAspLeuSerGlnInValPro 1738
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 TCGCTGGCACTCAACTACGCTGCGGGTCCCGGCATCATCGACCTGTTCCAAGTGCCA 481

Qy  1739 HisLeuProValLeuValProThrProGlyThrProAlaThrAlaMetAspArgArgLeu 1758
         |||||||||||||||||||||||||||||||||||||||||||
Db   482 CACCTGCCTGTGCTCGTGCCCCCGACACCAGGCACCCCAGCCACCGCCATGGACCGNCTT 541

Qy  1759 AlaTyrLeuProThrAlaProGlnProPheSerSerArgHisSerSerProLeuSer 1778
         ||||||||||||||||||||||||||||||||||||||||||
Db   542 GCCTACCTNCCCACCGCCCCACAGCCCTTCAGCAGCCGCCACAGCAGCTCCCCACTCTCC 601

Qy  1779 ProGlyGlyProThrHisLeuThrLysProThrThrSerSerGluArgGluArg 1798
         ||||||||||||||||||||||||||||||||||||||||||||
Db   602 CCAGGAGGTCCAACACACTTGAC-AAACCAACACCACGTTCCTCGTCCGAGCGGGGAGCGA 660

Qy  1799 Asp-ArgArgGluAspArgAspArgAspProGlyAspArgGluLysSerIleLeuThrSerTh 1818
         ||||:||:::||||||||||||||||||||
Db   661 GACCCGGATCGAGAGCGGGACCGGAGTCCATTCTCACGTCCAC 720

Qy  1818 rThrThrValGluHisAlaProIleTrpArgProGlyThrGlnSerSerGlySerSe 1838
         |||||||||||||||||||||||||||||||||||||||||
Db   721 CACGACGGTGGAGCACCACGCACCCATCTGGAGACCCTGTACAGACANAGCANNCGCAGCAG 780

Qy  1838 rGlySerSerGlyGlyGlyGlyGlySerSer---SerArgProAlaSerHisSerHisSerHisAl 1857
         ||||:::||||||||||||||||||:::
Db   781 CGGAGCCACGGGGGGGTGGGGGGGCCNTCCGTTTTCTTTCCT- 823

Qy  1857 aHisGlnHisSerProIleSerProArgThrGlnAspAlaLeuGlnArgProSerVa 1877
         ||||:::||||||||||||||
Db   824 ------------TTTCCCGTT---CCCCGC----------- 838

Qy  1877 lLeuHisAsnThrGlyMetLysGlyIleIleThrAlaValGluProSerLysProThrVa 1897
         ||||||||||||||||||||||||||||||||||
Db   839 --------------TTCCTCNCTGTTTNTCTGCCCGTGCCCTCTCGT 873

Qy  1897 lLeuArgSerThrSerSerProValArgProAlaAlaThrPheProProAlaTh 1917
         |||          :::::|||           :|||
Db   874 TCTC---------TTTGTTCCCCTCCCTTTTCTCCCCCTCCCCTTTTCGCGGTTCTTC 924

Qy  1917 rHisCysProLeu 1921
         |||||||
Db   925 TTCTTCGCCCCTT 937

RESULT 36
BU172348
LOCUS       BU172348           898 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7970482 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6164778
            5', mRNA sequence.
ACCESSION   BU172348
VERSION     BU172348.1  GI:22686332
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 898)
```

```
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC/DCTD/DTP
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM13522 row: g  column: 19
           High quality sequence stop: 593.
FEATURES             Location/Qualifiers
     source          1..898
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6164778"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb.  Library constructed by Life
                     Technologies."
ORIGIN

Alignment Scores:
Pred. No.:               8.29e-45              Length:           898
Score:                   1305.50               Matches:          265
Percent Similarity:      86.32%                Conservative:     0
Best Local Similarity:   86.32%                Mismatches:       17
Query Match:             9.88%                 Indels:           25
DB:                      13                    Gaps:             6
```

```
US-09-522-753-5 (1-2517) x BU172348 (1-898)

Qy   1462 GlyAlaSerThrThr-GlySerLysLysHisAspValArgSerLeuIleGlySerProGl 1481
          ||||||||||||||||  |||||||||||||||||||||| |||||||||||||||||||||
Db      3 GGGCGGCGTCCACCTGGCGTCCAAAAAGGACGACGTACGGCTCCCTCATCGGCAGCGCCGG 62

Qy   1481 yArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluAr 1501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 CCGGACGTTCCCACCCGTGCACCCGCTGGATGTGATGGCCGACGCCCGGGCACTGGAACG 122

Qy   1501 gAlaCysTyrGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySe 1521
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 TGCCTGCTACGAGAGAGCCTGAAGAGCCGGCCGAGCAGCGCCAGCTCCGGCGGGGTC 182

Qy   1521 rIleAlaArgGlyAlaProValIleValProGluGluGlyLysProArgGlnSerProLe 1541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 CATTGCGCGGCGGCGCCCCCGGTCATTGTGCCTGAGCTGGGTAAGCCGCGGCAGAGCCCCT 242

Qy   1541 uThrTyrGluAspHisGlyAlaProPheAlaGlyHisHisLeuProArgGlySerProValTh 1561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 GACCTATGAGGACCACGGGGCACCCTTTGCCGGCCACCTCCCACGAGGTTCGCCGTGAC 302

Qy   1561 rMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGl 1581
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 CACGCGGGAGCCCACGCCGCGGCCCTGCAGGAGGGCACCCTTTCGTCCAGCAAGGCATCCCA 362

Qy   1581 nAspArgLysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValPr 1601
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 GGACCGAAAGCTGACGTCGACGTGCGACGCCTCGTGAGATCGGCAGTCCCCGGCACCGTGCC 422

Qy   1601 oGluHisProHisProIleSerProTyrGluHisGlyLeuArgGlyValSerGlyVa 1621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 CGGAGCACCACCCCACACCCCATCTCGCCCTATGAGCACGGCCTGCGGGGCGTGGCGGT 482

Qy   1621 lAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIl 1641
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    483 GGACCTGTATCGCAGCCACATCCCCCTGGCCTTCGACCCCACCTCCATACCCCGGGGCAT 542

Qy   1641 eProLeuAspAlaAlaAlaAlaTyrLeuProArgHisLeuAlaProAsnProThrTy 1661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 CCCTCTGGACGCAGCCGCTGCCTACTACCTGCCCCGACACCTGGCCCCCAACCCCACCTA 602

Qy   1661 rProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAs 1681
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 CCCGCACCTGTACCCACCCTACCTCATCCGCGGCTACCCGACACGGCGGCGCTGGAGAA 662

Qy   1681 nArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaTh 1701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 ACGGCAGACCATCATCAATGACTACATCACCTCGCAGCAGATGCACCACAACACGGCCAC 722

Qy   1701 rAlaMetAla---GlnArgAlaAspMetLeuArgGlyLeuSer---ProArgGluSerSe 1719
          ||||||        ||||||||||||||||||||||||        |||||||||||
Db    723 CGCCGCCANCGAGCTGGATATGGCTGAGGGCCCTCTCCCCCCGCGAGTCCTC 782

Qy   1719 rLeu-AlaLeuAsnTyrAlaAla--------------------------GlyProA 1729
          ||||   |||||||||||||                              |||
Db    783 CCTGGCACTCAACTACCCCGTGCCCGGTCCCCCGAGGGCATTCATCGAACCGGTCCCCAA 842

Qy   1729 rgGlyTyrIleIleAspLeuSerGlnValProHisLeuProValLeuVal--------ProP 1746
          |||||||||||||||||||||||||||||||||||||||||          ||||
Db    843 GGGGC----------------------CCCCAC--CCTGCCCTGGTGCTTCGGGGCCCCC 878

Qy   1746 roThrProGlyThrPro 1751
          |||||||||||||||||
Db    879 CCAACCCCCAGGGGCCC 895
```

```
RESULT 37
BM560255
LOCUS       BM560255                1045 bp    mRNA    linear    EST 20-FEB-2002
DEFINITION  AGENCOURT_6563840 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5744688
            5', mRNA sequence.
ACCESSION   BM560255
VERSION     BM560255.1  GI:18804529
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1045)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12766  row: h  column: 01
            High quality sequence start: 18
            High quality sequence stop: 616.
FEATURES             Location/Qualifiers
     source          1..1045
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5744688"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_119"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27. Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning).  Average insert size 1.3 kb, insert size range
                     0.9-3 kb. Library is normalized and enriched for
```

```
Db    846  GAGGGCGGGCCGCTCCATCCATGGAGATCCCCGCTCCAGGAAACTGCGGACCACCCC  905
Qy   1441  euAlaProArgProLeu 1446
                ||||||
Db    906  CCCAACCTGCCCCCCTG 922
```

RESULT 38

```
BQ691710
LOCUS       BQ691710                875 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8046381 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6209388
            5', mRNA sequence.
ACCESSION   BQ691710
VERSION     BQ691710.1  GI:21817026
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 875)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM23368 row: j   column: 13
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1..875
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6209388"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
                     Site 2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hu in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN
```

Alignment Scores:
```
Pred. No.:              875
Score:                  1274.00
Percent Similarity:     89.90%
Best Local Similarity:  88.85%
Query Match:            9.64%
DB:                     13
```

Conservative: 3
Mismatches: 17
Indels: 12
Gaps: 1

US-09-522-753-5 (1-2517) x BQ691710 (1-875)

```
Qy  1334  AlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHisIleArgGly 1353
          |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db     1  GCCATCCCCGCCGGAGCGACACAGCCCCACCACCCTCAAAGAGCAGCACCACCATCCGCGG 60

Qy  1354  SerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyrLeuArgArg 1373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TCCATCACACAAGGGATCCCTCGGTCCTACGTGGAGGCACAGGAGGACTACCTGCGTCGG 120

Qy  1374  GluAlaLysLeuLeuLysArgGluGlyThrProProProSerProSerArgAspAspLeu 1393
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GAGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGCCCCCCAGCCCTTCACGGGACCTG 180
```

```
Qy  1146  GlyPro-------ValThrMetGlyLeuProMetAspProLysLysLeuAla 1162
          |||||           :::||||||||||||||||||||||||||||||||
Db     6  GGTCCGGAATTCCTTGATCCATGGGGCTGCCCCTGCCCCATGGACCCCAAAAAGTTGGCA 65

Qy  1163  ProPheSerGlyValLysGlnGluGlnLeuSerProArgGlyGlnAlaGlyProProGlu 1182
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66  CCCTTCAGCGGAGTGAAGCAGGAGCAGCTGTCCCCACGGGGCCAGGCTGGGCCACCGGAG 125

Qy  1183  SerLeuGlyValProThrAlaGlnGluAlaSerValLeuArgGlyThrAlaLeuGlySer 1202
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126  AGCCTGGGGGTGCCCACAGCCCAGGAGGCGTCCGTGCTGAGAGGGACAGCTCTGGGCTCA 185

Qy  1203  ValProGlyGlySerIleThrLysGlyIleLeuProSerThrArgValProSerSerAla 1222
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186  GTTCCGGGCGGGAGCATCACCAAAGGCATTCCAGCACCACGGGTGCCCTCGGACAGCGCC 245

Qy  1223  IleThrThrTyrArgGlySerIleThrHisGlyThrProAlaAspValLeuTyrLysGlyThr 1242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246  ATCACATACCGCGGCTCCATCACCCACGGCACGCCAGCTGACGTCCTGTACAAGGGGCACC 305

Qy  1243  IleThrArgIleIleGlyIleleGlyTyrGluSerProSerArgGlyArgAspArgGlyArgGluAspSer 1262
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306  ATCACCAGGATCATCGGGGAGGACAGCCCGAGTCGCTTGGACCGGGCCGGGAGGGAGGACAGC 365

Qy  1263  LeuProLysGlyHisValIleTyrGluGlyLysLysLysGlyHisValLeuSerTyrGlyGly 1282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   366  CTGCCCAAGGGCCACGTCATCTACGAAGGCAAGAAGGGCCACGTCTTGTCCTATGAGGGT 425

Qy  1283  GlyMetSerValThrGlnCysSerSerGluSerGlySerGlyProProHis 1302
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   426  GGCATGTCTGTGACCCAGTGCTCCAAGGAGAGCGGCAGAGCAGCTCAGGACCCCCCAT 485

Qy  1303  GluThrAlaAlaProLysArgThrTyrAspMetMetGluGlyArgValGlyArgValGlyArgAlaIle 1322
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   486  GAGACGGCGGCCCCCAAGCGCCACCTATGACATGATGGAGGGCCGGGTGGGCAGAGCCATC 545

Qy  1323  SerSerAlaSerIleGluGlyLeuMetGlyArgAlaIleLeuProProGluArgHisSerPro 1342
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   546  TCCTCAGCCAGCATCGAAGGTCTCATGGGCCGTGCCATCCCCGAGCGGCACAGCCCC 605

Qy  1343  HisHisLeuLysGluGlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSer 1362
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   606  CACCACCTCAAGGAGCAGCACCACATCCGCGGGTCCATCACACAAGGGATCCCTCGGATC 665

Qy  1363  TyrValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeu-LeuLysArgGluGl 1382
          ||||||||||||||||||||||||||||||||||||||||||||
Db   666  TACGTGGAGGCACAGGAGGACTACCTGCGTCGGGAGGCCAAGCTCCTAAAGCGGGAGGG 725

Qy  1382  yThrProProProSerProSerArgAspAspLeuTyrLysThrGlnAlaLe 1402
          |||  ||||||||||||||||||||||||||||||||||  |||||
Db   726  NACGNCTCCGACCCCATGGCCCCTGACGTCCATCACAAAAACGCAGGCCCC 785

Qy  1402  uGlyProLeu---LysLeuLys-ProAlaHisGluGlyLeuValAlaThrValLysGluA 1421
          |||||||    ||||||||  :|||::||||||||||||||||||||||||
Db   786  TGGGGCCCCTGAAACCTGAAACCCGGCCCATGAAGGGCCTGGTGGCCACGGTGGAAG 845

Qy  1421  laGlyArgSerIleHisGlyIleProArgGluGluLeuProGlyLeuProLeuProL 1441
          |||:::   |||          :::::||||||||||||||
Db   821  GAGGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGCCCCCCAGCCCTTCACGGGACCTG 180
```

ORIGIN

Alignment Scores:
```
Pred. No.:              1045
Score:                  1305.00              Matches:       267
Percent Similarity:     87.95%               Conservative:  3
Best Local Similarity:  86.97%               Mismatches:    31
Query Match:            9.88%                Indels:        6
DB:                     12                   Gaps:          2
```

US-09-522-753-5 (1-2517) x BM560255 (1-1045)

```
FEATURES             Location/Qualifiers
     source          1..915
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30512556"
                     /tissue_type="Brain glioblastoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_192"
                     /note="Vector: pExpress1; Site_1: SmaI; Site_2: NotI; The
                     library was constructed by reverse transcription of 1 ug
                     mRNA using the oligo dT primer GCGGCCGCCC(T)20 and an
                     RNaseH + MMLV reverse transcriptase. Second strand
                     synthesis was carried out by standard methods. The cDNA
                     was size selected by agarose gel for > 1.2 kb, digested
                     with Not I and directionally cloned into the vector
                     Express-1 at the SmaI/NotI sites. DNA from the primary
                     library was used for in vitro transcription from the T7
                     promoter to produce biotinylated RNA transcripts. These
                     biotinylated transcripts, along with blocking oligos to
                     the poly-A, multiple cloning site and primer regions, were
                     hybridized with single stranded circles produced by
                     phagemid production from the primary library to a Cot
                     value of 10-20. Strepavidin/phenol extraction was utilized
                     to remove DNA:RNA hybrids leaving un-hybridized single
                     stranded circles which were repaired by primer extension
                     and transformed back into E. coli resulting in the
                     normalized library. Average insert size 2.0 kb. 3'
                     linker/adaptor sequence GCGGCCGCCC(T)20. This library was
                     constructed by Agencourt Bioscience."
```

```
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM573  row: a  column: 13
High quality sequence stop: 627.
```

```
ORIGIN

Alignment Scores:
Pred. No.:                             Length:               915
CF272442                  5.1e-43      Matches:              265
Score:                    1262.50      Conservative:           9
Percent Similarity:        77.40%      Mismatches:            17
Best Local Similarity:     74.86%      Indels:                63
Query Match:                9.55%      Gaps:                   4
DB:                            14

US-09-522-753-5 (1-2517) x CF272442 (1-915)

Qy    2149  ArgHisHisProGlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGly  2168
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3  CGGCACCACCCCACAGCAGCTCAGGGACCCTGCCCGCCCCCTCACTCCTTCCTCCTGG      62

Qy    2169  AlaSerCysProValLeuAspLeuArgArgProSerAspLeuTyrLeuProProPro  2188
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63  GCCAGCTGCCCCGTCCTGGACCTCCGGAGACCCTCAGTGACCTCTACCTCCCGCCCCCG  122

Qy    2189  AspHisGlyAlaProAlaArgGlyGlySerProHisSerGluGlyGlyLyGlyArgSerProGlu  2208
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123  GACCATGGTGCCCCGGCCCGTGGCTCCCCCCACAGCGAAGGGGCCAAGAGGTCTCCAGAG  182

Qy    2209  ProAsnLysThrSerValLeuGlyValGlyValGlyIleGluProValSerProPro  2228
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     183  CCAAACAAGAGACGTCGGTCTTGGGTGGTGGTGGTGATTGAACCTGTGTCCCCACCG  242

Qy    2229  GluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgArgAsp  2248
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     243  GAGGGCATGACGGAGCCAGGGCACTCCCGGAGTGTGTGCTGTACCCGGAT  302

Qy    2249  GlyGluGlnThrGlnProSerArgMetGlySerLysSerLysSerSerProGlyAsnThrSerGlnPro  2268
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     303  GGGAACAGCAGACGGAGCCCAGGATGGGCTCCAGCAAGTCTCCAGGCAACACCAGCCAGCCG  362

Qy    2269  ProAlaPheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGln  2288
```

```
Qy    1394  ThrGluAlaTyrLysThrGlnAlaAlaLeuGlyProLeuProLeuLysProAlaHisGluGly  1413
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181  ACCGAGGCCTACAAGACGCAGGCCCTGGGCCCCCTGGAAGCTGAAGCCGGCCATGAGGGC  240

Qy    1414  LeuValAlaThrValLysGluAlaGlyAlaGlyArgSerIleHisGluIleProArgGluLeu  1433
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241  CTGGTGGCCACGGTGAAGGAGGCGGGCCGGCCTCCATCCATGAGATCCCGCGGAGGAGCTG  300

Qy    1434  ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGlyGlySerIleThrThrGln  1453
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301  CGGCACACCCCGAGCTGCCCCTGGCCCCGCCGCCGCTCAAGGAGGGGCTCCATCACGCAG  360

Qy    1454  GlyThrProLeuLysTyrTyrAspThrGlyAlaSerThrThrGlyLysLysGlyHisAspVal  1473
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361  GGCACCCCGCTCAAGTACTACGACACCGGCGCGTCCACCACTGGCTCCAAAAGACGACGTA  420

Qy    1474  ArgSerLeuIleGlySerProGlyArgThrPheProProValHisSerProLeuAspValMet  1493
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421  CGCTCCCTCATCGGCAGCCCCGGCCGGACGTTCCCACCCGTGCACCCGCTGGATGTGATG  480

Qy    1494  AlaAspProAlaAlaLeuGluArgAlaCysTyrGluGluSerLeuLysSerArgProGly  1513
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481  GCCGACCCGGCGGCACTGGAACGTGCCTGCTACGAGGAGAGCCTGAAGAGCCGGCCAGGG  540

Qy    1514  ThrAla-SerSerSerGlyGlySerIleAlaAlaArgGlyAlaProValIleValProGluLe  1533
            |||||    |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541  ACCGCCAGNCAGCTCGGGGGCTCCATTGCCGCGCGGGGCGCCCGGTCATTGTGCCTGAGCT  600

Qy    1533  uGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHi  1553
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601  GGGTAAGCCGCGGCAGCGCCCCTGACCTATGAGGACCACGGGGCACCCTTTGCCGGCCA  660

Qy    1553  sLeuProArgGlySerProValThrMetArgGluProThrProArg-LeuGlnGluGlyLys  1573
            |||||||||||||||| |||||||||||||||||||  ||||||
Db     661  CCTCCCACGGAGGTTGCCCGTGACCACGGGGGGAGCCCACGCGGCCCCTGCAGGAGGAGGCA  720

Qy    1573  erLeuSerSer-SerLysLysAlaSerGlnAsp-ArgLysLeuThr----------  1586
            |||||||   |||::::|   |||:::
Db     721  GCCTTTCGTCCAGGCAAGGCATCCCAAGAACCGAAAGCTGACCGTCGACGCCCTCCGTGA  780

Qy    1587  ------SerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHi  1604
                  ||||||||||    |||:::
Db     781  GAATCGAAGGTCCCCGCCACAGGCAACACGGGCCCGGAAACACCACCCAAACCCCA  840

Qy    1604  sProHisProIle  1608
            |||:::|||
Db     841  TTCTTCCGCCCTA  853
```

```
RESULT 39
CF272442
LOCUS       CF272442             915 bp   mRNA    linear   EST 13-AUG-2003
DEFINITION  AGENCOURT 15178952 NIH_MGC_192 Homo sapiens cDNA clone
            IMAGE:30512556 5', mRNA sequence.
ACCESSION   CF272442
VERSION     CF272442.1  GI:33628354
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Agencourt
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
              High quality sequence start: 2
              High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1..1075
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5501184"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."

ORIGIN

Alignment Scores:
Pred. No.:              7.73e-43        Length:          1075
Score:                  1260.00         Matches:         236
Percent Similarity:     99.59%          Conservative:    8
Best Local Similarity:  96.33%          Mismatches:      1
Query Match:            9.53%           Indels:          0
DB:                     12              Gaps:            0

US-09-522-753-5 (1-2517) x BM461469 (1-1075)

Qy   2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnGluIleAsnLys 2292
          ||||:::::: ||  :::::::::::::::::::::::::::::::::::::::::::
Db      1 TCCCGGGTCTCGTCGACCAGCCGTCGTGGTCAAGTCCAAGAAGAAGATCAACAAG 60

Qy   2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCTGGACGGAG 120

Qy   2313 IlePheAsnMetProAlaIleThrGlyThrLeuMetThrTyrArgSerGlnAlaVal 2332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAAGCCAGGGTG 180

Qy   2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleLeuArgGlyAlaLeuMetGly 2352
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db    181 CAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGGT 240

Qy   2353 LysTyrArgAspHisLeuProAlaAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspSerAspHis 2372
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db    241 AAATATGACCAGTGGAAGAGTCCCCGCGCTCAGCGGCCAATGCTTTTAACCCTCTGAAT 300

Qy   2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspSerAspHis 2392
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCAGTGCCAGCCTGCCCGGCTATGCCGGTGCTGACGGACGGAGTGACCAC 360

Qy   2393 ThrLeuThrSerProGlyTyrGlyGlyLysGlyLysValSerGlyTyrProSerSerArg 2412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACACTCACCTCGGCCAGGTGGGCGGGGAAGGCCAAGGTCTCTGCAGACCCAGCAGCCGA 420

Qy   2413 LysAlaLysSerProAlaAlaProGlyLeuAlaSerGlyAspArgProSerValSerSer 2432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAGCCAAGTCCCCGGCCGCCCCGGGCCTGGCATCGGGGGACCGGCCACCCTGTCTCCTCA 480

Qy   2433 ValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAsp 2452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAACCGCGTGTGGGAGGAC 540

Qy   2453 ArgProSerSerAlaGlyTyrThrProPheProTyrAsnProLeuIleMetArgLeuGln 2472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGGCCCTCCTCGGCCGGGTACACCCCATTCCCCTACAACCCCCTGATCATGCGGCTGCAG 600

Qy   2473 AlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyProLeuProAla 2492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCGGGGGTCATGGCTTCACCTCCCCACCCGGGCCAGCCGGGCGGCGGCCCCCTGCCT 660

Qy   2493 GlyProHisHisAlaTrpAspGluProLysProLeuCysSerGlnTyrGluThr 2512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    363 CCAGCCTTCTTCAGCAAGCTGACCGAGCAACTCCCCCATGCCAGTCAAGAAGCAA 422
          ||||||||||||||||||||||||||||||||||||||||||||

Qy   2289 GluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGln 2308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 GAGATCAACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCAGCCAG 482

Qy   2309 ProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrLeuMetThrTyrArg 2328
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483 CCTGGGGACGGAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGA 542

Qy   2329 SerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleLeuArgGlyLys 2348
          ||||||||||||||||||||||||||||||||||  |||  ||| |||||||
Db    543 AGCCAGGGCGGTGCAGGAACATGCCAGCACCAACATGGGCTGGAGGCCATAATTAGAAAAG 602

Qy   2349 AlaLeuMetGlyLysLysTyrArgAspHisLeuProGluSerProProLeuSerProAlaAsnAlaPhe 2368
          ||||||||
Db    603 GCACTCATG---------------- 611

Qy   2369 AsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyGly 2388
          ||||||||||||||||||||||||||||||||||
Db    611 AC ------------------------ 611

Qy   2389 ArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyGlyLysAlaLysSerValGlyArg 2408
          ||||||||||||||||||||||||
Db    612 --------------GGTGGCGCGGGAAGGCCAAGGTCTCTGCGCAGA 644

Qy   2409 ProSerSerArgLysAlaLysSerProAlaProGly-LeuAlaSerGlyAspArgPro-P 2428
          ||||||||||||||||||||||||||||||||||||||   |||
Db    645 CCCAGCAGCCGAAAAGCCAAGTCCCCGGCCCGGCCCGGCATCTGGGACCGGCCACC 704

Qy   2428 roSerValSer-SerValHisSer-GluGlyAspCysAsnArg-----ArgThrProLe 2445
          ||||||||||||||||||  |||||||||  ||||||||||||  |||
Db    705 CCTCTGTCTCCTCCAGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCAAC 764

Qy   2445 uThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrProPheProTyrAs 2465
          ||||||||||||  |||:||  ||  |:||
Db    765 CAACCGCCGTGTGGGAGGACAGGCCCTTCGTCCGGCCAGGTTCCACCCGCCATTTCC 824

Qy   2465 nPro----------LeuIleMetArgLeuGln--AlaGlyValMet----------- 2476
          ||||                :::::
Db    825 CCCTTCAAACCCCCCCCGGATCATGGGCCTGCAGGGGCGGCGGGGTGGTCAATGGGNTT 884

Qy   2477 --AlaSerProProProProGlyPro 2485
          :::||||||||||| ||| |
Db    885 TTCCCCACACCCCCCCCACCGGGGGGCCCT 914

RESULT 40
BM461469
LOCUS       BM461469                1075 bp    mRNA     linear    EST 05-FEB-2002
DEFINITION  AGENCOURT_6420985 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5501184
            5', mRNA sequence.
ACCESSION   BM461469
VERSION     BM461469.1  GI:18510509
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1075)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov at:
            Plate: LLAM12137  row: f  column: 01
```

Db   661 GGCCCCCACCACGCCTGGGACGAGGAGCCCAAGCCACTGCCTCGTCCGCAGTACGAGACA 720

Qy   2513 LeuSerAspSerGlu 2517
Db   721 CTCTCCGACAGCGAG 735

RESULT 41
BU164114
LOCUS   BU164114   928 bp   mRNA   linear   EST 04-SEP-2002
DEFINITION AGENCOURT_7939082 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6008388
            5', mRNA sequence.
ACCESSION   BU164114
VERSION   BU164114.1 GI:22678066
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 928)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13193 row: c column: 13
            High quality sequence stop: 620.
FEATURES   Location/Qualifiers
     source   1..928
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6008388"
                     /tissue_type="large cell carcinoma"
                     /clone_lib="NIH_MGC_68"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."

ORIGIN
Alignment Scores:
Pred. No.:   Length: 928
Score: 9.18e-43   Matches: 251
Percent Similarity: 85.95%   Conservative: 6
Best Local Similarity: 83.95%   Mismatches: 13
Query Match: 9.51%   Indels: 29
DB: 13   Gaps: 3

US-09-522-753-5 (1-2517) x BU164114 (1-928)

Qy   1847 SerSerSerArgProAlaSerHisSerHisAlaHisGlnHisSerProIleSerProArg 1866
Db   2 AGCAGCAGCAGGCCCCGCGCCTCCCACTCCCATGCCCACCACCAGCACTCGCCCATCTCCCCTCGG 61

Qy   1867 ThrGlnAspAlaLeuLeuGlnLeuArgProSerValLeuHisAsnThrGlyMetLysGlyIle 1886
Db   62 ACCCAGGATGCCCTCCTCCAGCAGACGACCCCAGTGTGCTTCACAACACAGGCATGAAGGGTATC 121

Qy   1887 IleThrAlaValGluProSerLysProThrValLeuArgSerThrSerThrSerPro 1906
Db   122 ATCACCGCTGTGGAGCCCAGCAAGCCCACGGTCCTGAGGTCCACCTCCACCTCCACCC 181

Qy   1907 ValArgProAlaAlaThrPheProProAlaThrHisCysProLeuGlyTyrThrLeuAsp 1926
Db   182 GTTCGCCCCGGCTGCCGCCACATTCCCACCTGCCACCACTGCCACCCTCGAT 241

Qy   1927 GlyValTyrProThrLeuMetGluProLysGluAlaProArgValAla 1946
Db   242 GGGGTCTACCCTACCCTCATGGAGCCCGTCTTGCTGCCAAGAGAGGCCCCCGGGTCGCC 301

Qy   1947 ArgProGluArgProAlaAspThrArgThrGlyMetLysProProAlaArg 1966
Db   302 CGGGCCAGAGACGCGGCCCGACGACCCGGCCATGCCCCTTCCTCGCCAAGCCCCCAAGCCCCGC 361

Qy   1967 SerGlyLeuGluProAlaSerSerProSerLysGlySerGluProArgProLeuValPro 1986
Db   362 TCCGGGCTGGAGCCCGCCTCCTCCCCAGCAAGGGCTCGGAGCCCCGGCCCCTAGTGCCT 421

Qy   1987 ProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHis 2006
Db   422 CCTGTCTCTGGCCACGCACCATCGCCCGCACCCCTGGCAAGAACCTCGCACCTCACCAC 481

Qy   2007 AlaSerProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThr 2026
Db   482 GCCAGCCCGGACCCCGCGCCACCTGCCTCGGCCTCGGACCCGCACCGGGAAAAGACT 541

Qy   2027 GlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHisGlySer 2046
Db   542 CAAAGTAAACCCTTTTCATCCAGGAACTGGAACTCCGTTCTCTGGTTACCACGGCAGC 601

Qy   2047 SerTyrSerProGlyGlyValGluProValSerProSerLeuThrHis 2066
Db   602 AGCTACAGCCCGGAAGGGGTGGAGCCCGTCAGCCCTGTGACCCCCAC 661

Qy   2067 AspLysGlyLeuProLysHisLeuGluLeuAspLysSerHisLeu---GluGlyGlu 2085
Db   662 GACAAGGGGCTCCCCAAGCACCTGGAAGAGCTCGACCACCTGGAGAAGGGGGAGC 721

Qy   2086 LeuArgProLysGlnPro-GlyProVal-LysLeuGlyGlyGluAlaAlaHisLeuProH 2105
Db   722 TGCGGCCCCAAGCAGCCAGGGCCCGTGGAAGCTTGGGGGGGAGGGGCC---------- 771

Qy   2105 isLeuArgProGluSerGlnProSerSer------------------------------ 2117
Db   772 ------------------------GCCCAACCTTCCACAACCTGGGGGGGCCGTCTTCTAAAAGT 811

Qy   2118 ------------------ProLeuLeuGlnThrAlaProGlyVal 2126
Db   812 ATTCCCCTTGGAGGGGGCCCAAAGGCCTCTAAGCGGGTAACCCCAGGGCTT 864

RESULT 42
BX397973
LOCUS   BX397973   994 bp   mRNA   linear   EST 13-MAY-2003
DEFINITION BX397973 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI051YA15 5-PRIME, mRNA sequence.
ACCESSION   BX397973
VERSION   BX397973.1 GI:30629329
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 994)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6282.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI051AA08QP1&cluster=6282.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ inVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI051AA08QP1.

```
Db   771 GCCGAGGGGGGGCGCTCAAGCAGCAGARAAGAAGGAGGGGGGGGAGGGGCAGGGCCACCACAGCC 830
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   907 LysSerSerGlyAlaProGlnAspSerAspSerSerAlaThrCysSerAlaAspGluVal 926
         ||||||||||||||||||||||||||||||||||||||||||||||||||     |||
Db   831 AARAGCTCCGGCGCCGCCCCCAGGACAGCAGCGACTCCAGTGCTACCTGCAGTGM-GACGAGGTG 889

Qy   927 AspGluAlaGluGluGlyGlyAspLysAsnArgLeuLeuSerProArgLeuLeuThr 946
         |||||||||||||||||||||||||||||||||||||||||||||     :::
Db   890 GATGAGGCGGAGGSSGGCGACAAAAA---CSGCTGCTGTC-CCAAGGCCAGCT--YCTCAC 942

Qy   947 ProThrGlyAspProArgAlaAsnAlaSer 956
         |||||||||      |||||
Db   943 CCGACTGCG---ACCCGGGCAAATCBTACC 969
```

RESULT 43
BM471347
LOCUS   BM471347   1066 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION AGENCOURT_6463899 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5517786
            5', mRNA sequence.
ACCESSION   BM471347
VERSION   BM471347.1 GI:18520389
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1066)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12177 row: i column: 19
            High quality sequence stop: 462.
FEATURES   Location/Qualifiers
     source   1..1066
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5517786"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "
ORIGIN

Alignment Scores:
Pred. No.:   Length: 1066
Score: 1.42e-42   Matches: 293
Percent Similarity: 1253.50   Conservative: 15
Best Local Similarity: 80.42%   Mismatches: 44
Query Match: 76.50%   Indels: 35
DB: 9.49%   Gaps: 4
                 12

```
US-09-522-753-5 (1-2517) x BM471347 (1-1066)

Qy   408 LysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGln 427
         ||::|||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db     4 CAGTTCATCAACATGAACGGGCTTATGGCCGACCCCATGAAGGTGTACAAAGACCGCCAG 63

Qy   428 ValMetAsnMetTrpSerGlnGluLysGluLysGluLlThrPheArgGluLysPheMetGlnHis 447
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    64 GTCATGAACATGTGGAGTGAGCAGGAGAAGGAGAAGGAGACCTTCCGGGAGAAGTTCATGCAGCAT 123
```

Alignment Scores:
Pred. No.:
Score: 1.04e-42   Length: 994
Percent Similarity: 1256.00   Matches: 271
Best Local Similarity: 88.39%   Conservative: 3
Query Match: 87.42%   Mismatches: 33
DB: 9.50%   Indels: 8
                 13   Gaps: 3

```
US-09-522-753-5 (1-2517) x BX397973 (1-994)

Qy   647 SerGlnCysLysAsnPheTyrPheAsnTyrTyrLysArgGlnAsnLeuAspGluIleLeu 666
         ||::|:      ||||||||||||||||||||||||||||||||||||||||
Db    54 TCCCGGGATAAGAACTTCTACTTCAACTACACAAGAAGAGGCAGAACCTTCGATGAGATCTTG 113

Qy   667 GlnGlnHisLysLeuLysMetGluLysGluValArgAsnAlaArgLysLysLysLysLysAla 686
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114 CAGCAGCACAAGCTGAAGATGGAAGAGGAAGAGGAAGAGCGCGGAGAAGAAGAAGAGAAAGCG 173

Qy   687 ProAlaAlaSerGluGluAlaAlaPheProProValValGluAspGluMetGlu 706
         ||||||||||||||||||||||||||||||||||||
Db   174 CCGGCGGS-GCCAGCGAGGAGGCTGCATTCCCGCCGTGTGTGGAGGATGAGAGGATGGAG 232

Qy   707 AlaSerGlyValSerGlyAsnGluGluGlnGluMetValGluValAlaLeuHisAla 726
         ||||||||||||||||||||||||||||||||||||||||||
Db   233 GCGTCGGGCGTGAC-GGAAATGAGGAGGAGAGATGGTGGAGGGCTGAGCCTTACATGCC 291

Qy   727 SerGlyAsnGluValProArgGlyGlyCysSerGlyValProAlaThrValAsnAsnSerSer 746
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   292 TCTGGGAATGAGGTGCCCAGAGGGAATGCGAGTGGCCCAGCCACTGTCAACAACAGCTCA 351

Qy   747 AspThrGluSerIleProSerProHisThrGlyAlaAlaAlaLysAspThrGlyGlnAsnGly 766
         |||||||||||||||||||||||||||||||||||||||||||||
Db   352 GACACCGAGAGCATCCCCTCTCCTCACACTGGAGGCCGCCAAGGACACAGGGCAGAATGGG 411

Qy   767 ProLysProProAlaThrLeuGlyAlaAspGlyProProGlyProGlyThrProPro 786
         |||||||||||||||||||||||||||||
Db   412 CCCAAHTCYCCAGCCACCCCTGGCCCGCCGGKCCWYCCCCAGGGCCTTCYCCWYCCCACCA 471

Qy   787 ArgArgThrSerArgAlaProIleGluProThrProAlaSerGluAlaThrGlyAlaPro 806
         |||    ||||||||||||||||||||||||||||||||
Db   472 CCGGAGCGACATCCCBNYYCCCACKGAGCMMACCCCGGCCTTCGAGGCMACCGGHBCCCCT 531

Qy   807 ThrProProAlaProProSerProSerProSerAlaProProValValProLysGluGlu 826
         ||||||||||||||||||||||||||||||||||||||||||
Db   532 ACBCCCCACCACCARCACCCCCAWCGCCCTCTGCACCTCCTCCTGTGTCCCCAARGAGGAG 591

Qy   827 LysGluGluThrAlaLeuAlaAlaProProValGluGluGlyGluGluGlnInLysProPro 846
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   592 AAGGAGGAGGAGACCCCAGCCAGCGCGCCCCCAGTGGAGGAGGGGGAGGAGCAGAAGCCCCCC 651

Qy   847 AlaAlaGluGluLeuAlaValAspThrGlyValLysAlaGluProValLysSerGluCys 866
         ||||||||||||||||||||||||||||||||||
Db   652 GCGGGCTGAGGAGCTGGCAGTGGACACAGGGAGGCCCCCGTCAAGAGCGAGTGC 711

Qy   867 ThrGluGluAlaGluGluLysLysLysLysAlaGlyLysAspAlaAlaGlyAlaAlaThr 886
         |||||||||||||  ||||||||||||||||||||||||||||   |||
Db   712 ACGGAGGAAGCCGAGGAGAAGAAAGAAGCCGGCAAGGACGCGGCCGGCGCTGAGGC-ACG 770

Qy   887 AlaGluGlyAlaAlaLeuLysAlaGluLysLysGluGluGlyGlySerGlyArgAlaThrThrAla 906
```

ACCESSION   IMAGE:6844056 5', mRNA sequence.
VERSION   CB723733
             CB723733.1 GI:29780875
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 797)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
             National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
             http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)
             Seq primer: pYX-5.

FEATURES   Location/Qualifiers
     source   1..797
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6844056"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
                     Site 2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 797
Score: 1.14e-42   Matches: 239
Percent Similarity: 1252.50   Conservative: 9
Best Local Similarity: 93.23%   Mismatches: 16
Query Match: 89.85%   Indels: 2
DB: 9.48%   Gaps: 1
                    14

US-09-522-753-5 (1-2517) x CB723733 (1-797)

Qy   1480 ProGlyArgThrPheProProValHisProLeuAspValMetAlaAspValArgAlaAlaLeu 1499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 CCCGGCCGCGCCTTTCCTCGCCTGCCACCCCCTGGACATAATGGCTGACGGCCCGGCACTG 61

Qy   1500 GluArgAlaCysTyrGluSerLeuLysSerArgProGlyThrAlaSerSerSerGly 1519
          |||||||||||||||||||||||||||||||||||||||||||||||::::||
Db   62 GAGCGTGCCTGCTATGAAGAGAGTCTGAAGAGCCGGTCAGGGACCAGCAGTGGTGCAGGG 121

Qy   1520 GlySerIleAlaArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSer 1539
          ||||||||||||||||||||||||||||||||::::||||||||||||||||||
Db   122 GGCTCCATCACACGTGGGGCTCCAGTCGTCGTGCCTGCCTGAACTGGCAAGCCACGGCAAAGC 181

LOCUS   CB723733   797 bp   mRNA   linear   EST 09-JUL-2003
DEFINITION UI-M-FY0-cez-b-23-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone

RESULT 44
CB723733

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4473272"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_90"
/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
```

ORIGIN

Alignment Scores:

Pred. No.:	1087
Score:	2.03e-42	Matches:	1250.00
Percent Similarity:	75.65%	Conservative:	276
Best Local Similarity:	72.25%	Mismatches:	13
Query Match:	9.46%	Indels:	61
DB:	12	Gaps:	33
			4

US-09-522-753-5 (1-2517) x BG252257 (1-1087)

```
Qy  1091  ThrAlaArgProValLeuProArgProProThrIleSerAsnProProProLeuIleSer  1110
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2  ACTGCCCGGCCCGTCCTGCCGCGGCCACCAACCATCTCCAACCCGCCCCCCATCTCC      61

Qy  1111  SerAlaLysHisProSerValLeuGluArgProThrIleSerGlnGlyMetSer  1130
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    62  TCTGCCAAGCACACCCCAGCGTCCTCGAGAGGCAAATAGTGCCATCTCCCAAGGAAATGTCG  121

Qy  1131  ValGlnLeuHisValProTyrSerGlnHisHisAlaLysAlaProValGlyValThrMet  1150
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   122  GTCCAGCTCCACGTCCCGTACTCAGAGCATGCCAAGGCCCCGTGGGCCCTGTCACCATG  181

Qy  1151  GlyLeuProLeuProMetAspProLysLeuAlaProPheSerGlyValLysGlnGlu  1170
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   182  GGGCTGCCCCTGCCCATGGACCCCAAAAAGTCTGCACCCTTCAGCGGAGTGAAGCAGGAG  241

Qy  1171  GlnLeuSerProArgGlyGlnAlaGlyTyrProGluSerLeuGlyValProThrAlaGln  1190
          |||||||||||||||||||||||||||||||||||||||||||
Db   242  CAGCTGTCCCCACGGGGCCAGGCTGGCTACCCGGAGAGCCTGGGGGTGCCCACAGCCCAG  301

Qy  1191  GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys  1210
          |||||||||||||||||||||||||||||||||||||||||||||
Db   302  GAGGCGTCCGTGCTGAGAGGAGCAGCTCTGGGCTCAGTTCCGGGCGGAGCATCACCAAA  361

Qy  1211  GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleThr  1230
          |||||||||||||||||||||||||||||||||||||||||||
Db   362  GGCATTCCCAGCACACGGGTGCCCTCGGACAGCGCCATCACATACCGGGCTCCATCACC  421

Qy  1231  HisGlyThrProAlaAspValAspLeuThrArgIleIleGlyIleGluGlyGluAsp  1250
          |||||||||||||||||||||||||||||||||||||||||||
Db   422  CACGGCACGCCAGCTGACGTGGACCTCACAAGGGACCACCAGGATCATCGGCGAGGAC  481

Qy  1251  SerProSerArgLeuAspArgGlyArgGlu--AspSerLeuProLysGlyHisValIleTy  1270
          ||||||||||||||||||||||||||:::::
Db   482  AGCCCGAGTCGGCTTGGACCGCGGACAGGCGCCAAAGGGCCACGTCATCTA  541

Qy  1270  rGluGlyLysLysGlyHisValLeuSerTyrGlu-GlyGlyMetSerValThrGlnCysS  1290
          ||||||||:::|||||||||:|:||||||||:
Db   542  CGAAGGGCAAGAAGGGCCACGTCTTGTCCTATGACGGGGTGGCATGTCTGTGACCCAGTGCT  601

Qy  1290  erLysGluAspGlyArg-SerSerSerGlyProProHisGluThrAlaAlaProLysArg  1309
          |||||||||:|||||||||||||||||:|||||||||:||||||||
Db   602  CCAAGGAGGACGGCAGAAAGCAGCTCAGGACCCCCCATGAGACCCCAAGCGC  661

Qy  1310  ThrTyrAspMetMet-GluGlyValGlyArgValGly--ArgAlaIleSerSerAla--SerIle  1327
          ||||||||:|||::|||||||:|||||||:|::
Db   662  ACCTATGACATGATGCGAGGCCACGTGGGCACGAAGCCATCTCCTCAGCAAGCAATC  721

Qy  1328  GluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGlu  1347
```

```
Qy  1540  ProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro  1559
          |||||||||||||||||||||||||||||||||||
Db   182  CCACTGACTTACGAAGACCACGGGGCACCCTTCACCGGTCACCTGCCCAGGTGGCTCCCCT  241

Qy  1560  ValThrMetAlaGluProThrProArgLeuGlnGluGlySerLeuSerLysSerAla  1579
          ||||||||||||||||||||||||||||||||||||
Db   242  GTGACCACGGCAGAGGGGGAGCCCACGCCACGGCTTCAGGAAGGCAGCCTCCTATCCGACCAAGGCG  301

Qy  1580  SerGlnAsnArgLysLeuValThrSerProThrProArgGluIleAlaLysSerProHisSerThr  1599
          ||||||||||||||||||||:|||||||||||
Db   302  TCCCAGGAACCGGAAGCTGACATCTACACCCCGGAGATCGCCAAGTCCCACACACAGCACT  361

Qy  1600  ValProGluHisHisProHisSerProIleSerProTyrGluHisLeuLeuArgGlyValSer  1619
          ||||||||||||||||||||||||:::
Db   362  GTGCCCGAGCACCACCCTCACCCCATCTCCCCTATGAGCACCTGCTTCGCGGCGTGACT  421

Qy  1620  GlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg  1639
          |||||||||||||||||||||||||||||||||||
Db   422  GGTGTGGACCTGTACCGGTCACATCCCATTGGCCCTTTGACCCCACCTCCATACCCCGA  481

Qy  1640  GlyIleProLeuAsp--AlaAlaAlaAla-TyrTyrLeuProArgHisLeuAlaProAs  1658
          ||||||||:::
Db   482  GGGATCCCTCTGAAGCAGCAGCCGCAGCCCTACTACCTGCCCCGGCACTTGGCCCCCAG  541

Qy  1658  nProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAl  1678
          :||||||||||||||||||||||||||||||
Db   542  CCCCACCTACCCCACCTGTACCCACCTTACCTCATCCGGGCTACCCTGACACGGCGGGC  601

Qy  1678  aLeuGluAsnArgGlnThrIleIleGlnAspTyrrIleThrSerGlnMetHisHisAs  1698
          |||||||||||||||||||||||||||||||
Db   602  CCTGGAGAACCGCCAGACCATCATCAATGACTACATCACCTCCCAGAGATGCCACCACAA  661

Qy  1698  nThrAlaThrAlaMetAlaGlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSe  1718
          ||||||||||||||||||||||||||||||||
Db   662  CGCTGCCACCGCCATGGCCCAGCGTGCTGACATGCTGAGGGGTCTGTCNACGGAGAGTC  721

Qy  1718  rSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGlyIleIleGlyHisSerGlnValPr  1738
          |||||||||||||||
Db   722  CTCGCTGGNCCTCAATTATGCCGCTGGCCCAAGAGGCATTATCGACCTGTCCCCAAGTGC  781

Qy  1738  oHisLeuProValLeu  1743
          |||||||||||
Db   782  ACACCTGCCCGTGCTG  797
```

RESULT 45

BG252257

LOCUS   BG252257   1087 bp   mRNA   linear   EST 13-FEB-2001

DEFINITION 602365136F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4473272 5',
            mRNA sequence.

ACCESSION   BG252257

VERSION   BG252257.1   GI:12762073

KEYWORDS   EST.

SOURCE   Homo sapiens (human)

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 1087)

AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.

TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL   Unpublished (1999)

COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10295 row: h column: 09
            High quality sequence stop: 641.

FEATURES   Location/Qualifiers
     source   1..1087

```
Db    722  GAACGTCTCATGGGCCGGCC-ATCCCCGGCAGCGACACCATTCACAAGAC 780
                ||      |||||||||                    ::::::
                ||      |||||||||   ||||||

Qy   1348  GlnHisHisIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGln 1367
                |||      |||           |||||||      |||
Db    781  CACCACCAATTCCGGGGTCCATCACCCAGGAT---CCCGGGTCTTCGTGTGAGACCGGA 837

Qy   1368  GluAspTyrLeuArgArgArgAlaAlaLysLeuLysArgGluGlyThrProProProPro 1387
                ||||::         ::      ::
Db    838  AGA--GCAACTGATGGGAGGGAGGCACGATCCTAAGCGGCGGACACCCTCCCCACGCCA 894

Qy   1388  ProProSerArgAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeu 1407
                |||      ||:::
Db    895  CAGGCCCTCCATGACATATAAACAGCCCAAATAAACTACCCCAACTACACTC 954

Qy   1408  LysProAlaHisGluGlyLeuValAlaThrValLysGlyAlaAlaGlyArgSerIleHisGlu 1427
                |||||||||||||||
Db    955  TTA------------------------------------------------ 957

Qy   1428  IleProArgGluLeuArgHisThrProGluLeuProLeuAlaProArgProArgLeuLys 1447
                ::|||::     ||||||
Db    958  --------------CAAAAAAGGACCACTAATTACTCACCACTACTCACAGTCCTCTCCCG 1005

Qy   1448  GluGlySerIleThrGlnGlyThrProLeuLysTyrAspLeuLysTyrAspThrGlyAlaSerThr 1465
                :::      |||
Db   1006  CAACATCTCCGCTCACACACATTACCCTCTCATCATTCCACACTCACGCTACC 1059


Alignment Scores:
Pred. No.:            46
Score:               1243.00
Percent Similarity:  88.47%
Best Local Similarity: 85.67%
Query Match:         9.41%
DB:                              US-09-522-753-5 (1-2517) x BG831424 (1-958)

Length:              958
Matches:             275
Conservative:        9
Mismatches:          28
Indels:              15
Gaps:                5


Qy   1292  GlnAspGlyArgSerSerGly-ProProHisGluThrAlaAlaProLysArgThrTy 1311
                ||||||||||||||||||||       |||||||||||||||||||||||||||||
Db      2  GAGGACGGCAGGCAGCTCAGGACGCCCCATGAGACGGCCGCCCCAAGCGACCACCTA  61

Qy   1311  rAspMetMetGluGlyArgValGlyArgAlaIleSerSerAlaSerIleGluGlyLeuMe 1331
                ||||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db     62  TGACATGATGGAGGGCCGCGTGGGCAGAGGCCATCCTCCTCAGCCAGCATCGATGGTCTCAT 121

Qy   1331  tGlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGlnGlnHisHisIl 1351
                |||||||||       ||||||||||||||||||||||||||||||||
Db    122  GGGCCGGGC-ATCCCGCCGGACGACAGCCCCCACCACTCAAAGAGCAGCACCACAT 180

Qy   1351  eArgGlySerIleThrGlnGlyIleLeuProArgSerTyrValGluAlaGlnGluAspTyrLe 1371
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  CCGGGGTCCATCACACAAGGGATCCCTGGTCCTACGTGGAGGCACAGGAGGACTACCT 240

Qy   1371  uArgArgArgAlaAlaLysLeuLysArgGluGlyThrProProProProSerArg 1391
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GCGTCGGAGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGGCCCCTCACG 300

Qy   1391  gAspLeuThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysProAlaHi 1411
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GGACCTGACCGAGGCCTACAAGACCCAGGCCCTGGGCCCCCTGAAGCCGGCCCA 360

Qy   1411  sGluGlyLeu-ValAlaThrValLysGlyAlaAlaGlyArgSerIleHisGluIleProArg 1431
                |||||||   :::||||||||||||||||||||||||||||||||||||||||||||
Db    361  TGAGGGCCTGGATGGCCACGGTGAAGGAGGCGGCGCTCCATCCATGAGATCCCGCGCG 420

Qy   1431  luGluLeuArgHisThrProGluLeuProLeuAlaProArgProArgLeuLysGluGlySerI 1451
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AGGAGCTGCGGCACACGCCGGAGCTGCCCCTGGCCCCGAGGCCGCGGCTCAAGGAGGGCTCCA 480

Qy   1451  leThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrArgGlySerLysLysH 1471
                ||||||||||||||||||||||||||||::|||||||||||||||||||||||||
Db    481  TCACGCAGGGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCAGGGGCTCCAAAAAGC 540

Qy   1471  isAspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuA 1491
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  ACGACGTACGCTCCTCATCGGCAGCCCCGGCCGGACGTTCCACCCGGTGCACCCGCTGG 600

Qy   1491  spValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluLeuLysSerA 1511
                ||||||||||||||||||||||||||||||||||||||||||||::    ::
Db    601  ATGTGATGGCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGCTGAAGAGC 660

Qy   1511  rgProGlyThrAlaSerSerGlyGlySerIleAlaArgGlyAlaProValIleValP 1531
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CGGCAGGGACACCGGCCAGCAGCTCGGGGGGTCCATTGCGCGCGGCGCCCCCGTCATTGTGC 720

Qy   1531  roGlu-LeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPhe 1550
                ||||   |||||||||||||||||||||:::||||
Db    721  CTGAGCTTGGGAAGCCCCGTGAGCCCCT---AACTATGAGGACCACGGGGCACCCTTT 777

Qy   1550  AlaGlyHisLeuProArgGlySerProValThrMetArgGluProArgLeuGln 1570
                |||    |||      :::|||      |||
Db    778  GC-GGACACTC--CCACGAGGGTGCCCGTG-ACCACGGAGAGCCCACG---CGGCCTGCA 831

Qy   1570  GluGlySerLeuSerSerLysSerAlaSerGlnAspArgLysLeuThrSerThrProArg 1590
                ::|||      :::|||      :::
Db    832  GAAGGCAGCCT-TCGTCCAGCAAGA-TCCCGA---ACGAAGGTGACGGTGACGGCTCGT 886

Qy   1591  GluIleAlaLysSerProHisSerThrValProHisHisProIleSer 1609
                |||                |||    :::::
```

```
Db     887 GAGATCGCC--AGTCCCGACAGA------CCGGCGCGAGCACCAAGACCATTGGCT 934

RESULT 47
BG974253
LOCUS       BG974253              810 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602843992F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4979758 5',
            mRNA sequence.
ACCESSION   BG974253
VERSION     BG974253.1  GI:14361890
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 810)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10978 row: g  column: 23
            High quality sequence stop: 787.
FEATURES             Location/Qualifiers
     source          1..810
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4979758"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam4"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                     Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and Differentiation 7, 3-11 (1996)."

ORIGIN

Alignment Scores:
Pred. No.:
Score:                4.19e-42
Percent Similarity:   1239.00      Length:        810
Best Local Similarity: 92.45%      Matches:       244
Query Match:          92.08%       Conservative:    1
DB:                   9.38%        Mismatches:     17
                      12           Indels:          3
                                   Gaps:            2

US-09-522-753-5 (1-2517) x BG974253 (1-810)

Qy 2211 LysThrSerValLeuGlyGlyGlyGluAspGlyIleGluProValSerProProGluGly 2230
            |||||  |||  |||   |||||||||||||||||||||||||
Db    2 AAAACATCGGTCCTGGGCAGCAGCGAGGATGCCATTGAGCCTGTGTCCCCACCAGAGGGC 61

Qy 2231 MetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlu 2250
            |||||||||| |||||||    |:||||||  |||||   |||||||||||||||
Db   62 ATGACTGAGCCAGGACATGCTCGGAGCACTCGGCTACCCACCACTGCTGTATCGAGACGGGGAA 121

Qy 2251 GlnThrGlyProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla 2270
            |||  |||  |||  |||||||||||||||||||||||||||||||||
Db  122 CAGGGCGAGCCC--AGGATGGGCTCTAAGTCTCCAGGCAACACCAGCCAGCCCCCAGCC 178

Qy 2271 PhePheSerLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleIle 2290
            ||||||||||||||||||||||
```

```
Db    179 TTCTTCAGTAAGCTGACTGAGAGCAACTCCCCATGTGAAGTCGAAGAAGCAGGAGATC 238

RESULT 48
BG974253
Qy 2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
            |||||||||||||||||||||
Db    239 AACACAGAAACTCAACACCCAACACCCGGAACAGCCAGAATACAATATTGGCCAGCCTGGG 298

Qy 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
            ||||||||||||||||||||
Db    299 ACGGAAATCTTCAACATGCCCGCCATCACTGGAACAGGCCTTATGACCTGTAGAAGCCAG 358

Qy 2331 AlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleLeuArgLysAlaLeu 2350
            |||||||||||||||||||||
Db    359 GCGGTGCAAGAACACGCCAGCACCAACATGGGCTAGAGCGCCATTATTAGAAAAGGCACTC 418

Qy 2351 MetGlyLysTyrAspGlnTrpGluSerProP'roLeuSerAlaAsnAlaPheAsnPro 2370
            |||||||||||||||||||||
Db    419 ATGGGTAAAATATGATCAGTGGGAAGAGCCCCCCGGCCGGCCGGTTTTTAACCCT 478

Qy 2371 LeuAsnAlaSerAlaSerLeuPro---AlaAlaMetProIleThrAlaAlaAspGlyArg 2389
            ||||||||||||  |||||||||||
Db    479 CTGAATGCCAGGCCAGTCTGCCCGGTGCTCTGCTATGCCCATAACCACTGCTGACGACGG 538

Qy 2390 SerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgPro 2409
            ||||||||||||||||||||||
Db    539 AGTGACCACCGCACTCACCTCGCCAGGTGGAGGTGGGAAAGCCAAGGTCTCTGGCGAACCT 598

Qy 2410 SerSerArgGlyAlaAlaLysSerProAlaProGlyLeuAlaSerGlyAspArgProProSer 2429
            ||||||||||||||||||||
Db    599 AGCAGCCGAAAAGCCAAGTCGCCAGCACCAGGCCTAGCGTCCGGAGACCGACCCCCTTCT 658

Qy 2430 ValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgVal 2449
            ||||||||||||||||||||
Db    659 GTCTCCTCAGTACACTCAGAGAGGGGACTGCAATCGCCGAACAACCATTCAACACCGTGTG 718

Qy 2450 Trp-GluAspProSerAlaGlySerThrProPheProTyrAsnProLeuIleMe 2469
            |||||||||||||||||||||
Db    719 TGGGGAGGACCGNGCCTCATCTCCAGGGTCCACGCATTCCCTAAAACCCTTTGATTAT 778

Qy 2469 tArgLeuGlnAla 2473
            |||||||||||
Db    779 GAGGCTACAGGCC 791

RESULT 48
BQ944732
LOCUS       BQ944732              962 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10036107 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6481219
            5', mRNA sequence.
ACCESSION   BQ944732
VERSION     BQ944732.1  GI:22360210
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 962)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov.
            Plate: LLCM2662 row: d  column: 20
            High quality sequence start: 139
            High quality sequence stop: 597.
FEATURES             Location/Qualifiers
     source          1..962
                     /organism="Homo sapiens"
```

```
Qy   2447 AsnArgValTrpGluAspArgProSerAlaGlySerThr-ProPheProTyrAsnPr 2466
          |||||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db    583 AACCGCGTGTGGGAGGACAGGCCCTCGTCCGCAGTTCCACGCCCATTCCCTACAACCC 642

Qy   2466 oLeuIleMetArgLeuGlnAla-GlyValMetAlaSerPro-ProProProGly-LeuPr 2485
          ||||||||||||||||||||||  ||||||||||||||||  |||||||||||||  ||
Db    643 CCTGATCATGCGGCTGCAGGCGGGCGGGTGTCATGGCCTCCCCACCGGGCCCCTCC 702

Qy   2485 oAlaGlySerGlyProLeuAlaGlyProHisHisAlaTrpProAspGlu-GluProLysProL 2505
          |||||||||||||||||||||||||||||||||||||||||||  ||::|||||||||
Db    703 GCGGGCAGGCGGCCCCTCGCTGGCCCCCCACCAGCCTGGGACGAGGAGGAGCCCAGCCAC 762

Qy   2505 euLeu-CysSerGlnTyrGluThrLeu-SerAspSerGlu 2517
          ||||| ||||||||||||||||||| |||||||||||
Db    763 TGCTCCTGCTCCGCAATACGAGACCCTCCTCCGACAGCGAG 802

RESULT 49
BM811122
LOCUS        BM811122                    1113 bp    mRNA    linear    EST 05-MAR-2002
DEFINITION   AGENCOURT_6599082 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5465831
             5', mRNA sequence.
ACCESSION    BM811122
VERSION      BM811122.1  GI:19127945
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1113)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1965 row: d  column: 24
             High quality sequence stop: 523.
FEATURES             Location/Qualifiers
     source          1..1113
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5465831"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_41"
                     /note="Organ: sKin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:                              Length:                1113
Score:              1.22e-41            Matches:               259
Percent Similarity: 1231.50             Conservative:          14
Best Local Similarity: 77.12%           Mismatches:            43
Query Match:        73.16%              Indels:                38
DB:                 9.32%               Gaps:                  7
                    12

US-09-522-753-5 (1-2517) x BM811122 (1-1113)
```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6481219"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC 40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G).Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 962
Score: 7.42e-42   Matches: 258
Percent Similarity: 1235.00   Conservative: 1
Best Local Similarity: 81.45%   Mismatches: 6
Query Match: 81.13%   Indels: 53
DB: 9.35%   Gaps: 1
                    13

```
US-09-522-753-5 (1-2517) x BQ944732 (1-962)

Qy   2207 ProGluProAsnLysThrSerValLeuGlyGlyGlyAspGlyIleGluGlyProValSer 2226
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CCAGGACCAAACAAGACGTCCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCC 60

Qy   2227 ProProGluGlyMetThrGlyProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCACCGGAGGGGCATGACGGGACCGGGCCACTCCCGGAGGTGCTGTTACCCGCTGCTGTAC 120

Qy   2247 ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGGATGGGGAACAGACAGAGCCGAGCCCCGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGC 180

Qy   2267 GlnProProAlaPhePheSerLysSerLeuThrGluSerAsnSerAlaMetValLysSerLys 2286
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGCCGCCAGCCTTCTTCAGCAAGTCGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAG 240

Qy   2287 LysGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle 2306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGCAAGAGATCAACAAGAAGCTGAACACCCACACCCGGAATGAGCCTGAAATACAATATC 300

Qy   2307 SerGlnProGlyThrValPheAsnMetProAlaIleThrGlyThrGlyLeuMetThr 2326
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCAGCCTGGCACGGAGATCTTCAATATGCCCCCATCACCGGAACAGCCCTTATGACC 360

Qy   2327 TyrArgSerGlnAlaValGlnGluHisHisAlaSerThrAsnMetGlyLeuGluAlaIleIle 2346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TATAGAAGCCAGGCGGTGCAGGAACATGCCAGCACCAACATGGGCTGGAGGCCATAATT 420

Qy   2347 ArgLysAlaLeuMetGlyValLysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsn 2366
          ||||||||||||||||
Db    421 AGAAAGGCACTCATG----------------------------------- 435

Qy   2367 AlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAla 2386
          -------------------------------------------------- 435

Qy   2387 AspGlyArgSerAspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSer 2406
                          |||||||||||||||||||||
Db    436 -----------------GGTGGCGCGGCGGGAAGGCCAAGGTCTCT 462

Qy   2407 GlyArgProSerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyLysAspArg 2426
          |||||||||||||||||||||||||||||||||||||||||||
Db    463 GGCAGAGCCCAGCCAGCCGAAAAGCCAAGTCCCCGGCCTGGCATCTGGGGACCGG 522

Qy   2427 ProProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThr 2446
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 CCACCCCTCTGTCTCCCTCAGTGCACTCGGAGGGAGACTGCAACCCCGCAGCCGCTCACC 582
```

```
BU191146
LOCUS       BU191146              861 bp    mRNA    linear    EST 04-SEP-2002
DEFINITION  AGENCOURT 7975031 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6081811
            5', mRNA sequence.
ACCESSION   BU191146
VERSION     BU191146.1  GI:22705130
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 861)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
COMMENT     Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2309 row: f  column: 20
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1..861
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6081811"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:              1.34e-41            Length:          861
Score:                  1227.50             Matches:         245
Percent Similarity:     88.61%              Conservative:      4
Best Local Similarity:  87.19%              Mismatches:       27
Query Match:            9.29%               Indels:            6
DB:                     13                  Gaps:              2
```

```
US-09-522-753-5 (1-2517) x BU191146 (1-861)

Qy  1334 AlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisIleArgGly 1353
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GCCATCCCGCCGGAGCGACACAGCCCCACCACCACCTCAAAGAGCAGCACCATCCGCGGG 60

Qy  1354 SerIleThrGlnGlyIleProArgGlyTyrValGluAlaGlnGluAlaGlyTyrLeuArgArg 1373
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCCATCACACAAGGGATCCCTCGGTCCTACGTGGAGGCACAGGAGGACTACCTGCGTCGG 120

Qy  1374 GluAlaAlaLysLeuLeuLysArgGluGlyGlyThrProProProProSerArgAspLeu 1393
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GAGGCGGCCAAGCTCCTAAAGCGGGAGGGCACGCCTCCGCCCCCACCGGGACCTG 180

Qy  1394 ThrGluAlaTyrLysThrGlnAlaLeuGlyProLeuLysLeuLysProAlaHisGluGly 1413
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACGGAGGCCTACAAGACCCAGGCCCTGGGCCCCCTGAAGCTGAAGCCGGCCCATGAGGGC 240

Qy  1414 LeuValAlaThrValLysGluAlaGlyArgSerIleHisGlulleProArgGluGlyLeu 1433
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  1457 LeuLysTyrAspThrArgThrGlySerThrGlyLysLysHisAspArgSerLeu 1476
         ||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTCAAGTACGACACCAGGACGGGCGTCCACCACTGGCTCCAAAAGCACGACGTACGCTCCTC 60

Qy  1477 IleGlySerProGlyArgThrPheProProValHisProLeuAspAlaAspAla 1496
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ATCGGCAGCCCCGGACGTTCCCACCCGTCGATGTGATGTGCCGACGCGT 120

Qy  1497 ArgAlaLeuGluLeuArgAlaCysTyrGluLeuLysSerArgProGlyThrAlaSer 1516
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGGGCACTGGAACGTGCCTGCTACGAGGAGAGCCGGCCAGGACCGGCCAGC 180

Qy  1517 SerSerGlySerIleAlaAlaArgGlyAlaProValIleValProGluLeuGlyLysPro 1536
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGCTCGGGGGGCTCCATTGCCGCGCGGGGCCCCGGTCATTGTGCCGGCTGGGCAAGCCG 240

Qy  1537 ArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArg 1556
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CGGCAGAGCCCCCTAACCTATGAGGACCACGGGGCACCCTTTGCCGGCCACCTCCCACGA 300

Qy  1557 GlySerProValThrMetArgGluProThrProArgLeuGlnLeuGlyGlySerLeuSerSer 1576
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGTTCGGCCGGTGACCACGCGGGAGCCCACGCCGCGCCTGCAGGAGGGCAGCCTTTCGTCC 360

Qy  1577 SerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlulleAlaLysSerProPro 1596
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGCAAGGCATCCCAGGACCGAAAGCTGACGTCGACGCCTCGTGAGATCGCCAAGTCCCCG 420

Qy  1597 HisSerThrValProGluHisHisProIleSerProTyrArgGluHisLeuLeuArg 1616
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CACAGCACCGTGCCCGAGCACCACCCCATCTCGCCCTATCGGGAGCACCTGCTTCGG 480

Qy  1617 GlyValSerGlyValAspLeuTyrArgSerHisIleProLeu-AlaPheAspProThrSe 1636
         |||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   481 GGGGTGAGTGGCGTGGACCTGTATCGGCAGCCACATCCCCTGGGCCTTCGACCCCACCTC 540

Qy  1636 rIleProArgGlyIleProLeuAspAlaAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAl 1656
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CATACCCCGCGGCATCCCTCTGGACGCGGCCGCTGCCTACTACCTGCCCCGACACCTGGC 600

Qy  1656 aProAsnProThrTyrProHisLeuTyrProProTyrLeuLeuArgGlyTyrProAspTh 1676
         ||||||||||||||||||||||::::|||||:::|||||||||||||
Db   601 CCCCAACCCCACCTACCCGCACCTGTACCCCTACCTCCTCCGCGGCTACCCCGACAC 660

Qy  1676 rAla-AlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIle-ThrSerGlnGlnMet 1695
         |||| ||||||||||||||||::||| |||::|:|||||:|||
Db   661 GGCCGGCGGCTGGAGAACCCGCAAACCCTCCTCTTGACTTCCTTCACCTCGCAGCCGATG 720

Qy  1696 HisHisAsnThrAla------------ThrAlaMetAlaGlnArgAlaAlaAsp 1708
         |||||||||||          ||||         :::
Db   721 CACCACAAAGGCGGCCCCCCCCCCATGGGCCCACCCGAACTTGGGCCACACCC 777

Qy  1709 MetLeuArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyPro 1728
         :::||||||||||||||||||||||||||||||||||||||||||||||||::
Db   778 CTTCTCCCCCCCCCAGCCCCTCGTTGGGGACTTCCACCAACCCCTGCGCCGGCCCCCCC 837

Qy  1729 ArgGlyIleIleAspLeuSerGlnValProHisLeuValProProThrPro 1748
         ||||||||          :::
Db   838 CGAGGA-----------AGGGCCCCACAC--------CAACTTGGGCACACCC 873

Qy  1749 GlyThrProAlaThrAlaMetAspArgLeuAlaTyrLeuProThrAlaProGlnProPhe 1768
         ||||                  ||||:::|||
Db   874 GGTCAACCCCGCGGGACCCGATCCGGGGGGA----------GGAGCGGAGAACCCGTGG 924

Qy  1769 --------------------------------SerSerArgHisSerSerSer 1775
         ||||||                          :::::||||
Db   925 CAATGAAAAAGCGGCAACGGGGATATTATATAACCTGGCAAGAACCCATTTCTTCC 984

Qy  1776 ProLeu--------SerProGlyGlyProThrHis 1784
         |||:::                |||||||||||
Db   985 CCGATTTGATTTACTGATATGGGTGGCCCTACACAC 1020
```

RESULT 50

```
Db   241 CTGGTGGCCACGGTGAAGGAGGCGGGGCGCCGCTCCATCCATGAGATCCCGCGCGAGGAGCTG 300

                /db_xref="taxon:9606"
                /clone="IMAGE:4473271"
                /tissue type="adenocarcinoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_90"
                /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.7 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."

Qy  1434 ArgHisThrProGluLeuProLeuAlaProArgProLeuLysGluLysGlySerIleThrGln 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CGGCACACGCCCGAGCTGCCCCCTGGCCCCGCGCCGCTCAAGGAGGGCTCCATCACGCAG 360

ORIGIN
Qy  1454 GlyThrProLeuLysTyrAspThrGlyAlaSerThrThrGlySerLysLysHisAspVal 1473
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACTGGCTCCAAAAAGCACGACGTA 420

                Alignment Scores:
                Pred. No.:
Qy  1474 ArgSerLeuIleGlySerProGlyArgPheProProValHisProLeuAspValMet 1493         Score:                    1225.50
              |||||||||||||||||||||||||||||||||||||||||||||||||||          Percent Similarity:        90.24%
Db   421 CGCTCCCTCATCGGCAGCCCCGGCCGACGTTCCACCCGTGCACCCGCTGGATGTGATG 480         Best Local Similarity:     87.88%
                                                                                Query Match:               9.27%
                                                                                DB:                        12

Qy  1494 AlaAspAlaArgAlaLeuGluArgAlaCysTyrGluLysLeuLysSerArgProGly 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||     US-09-522-753-5 (1-2517) x BG252161 (1-876)
Db   481 GCCGACGCCCGGGCACTGGAACGTGCCTGCTACGAGGAGGCCTGCAAGGAGCCGCCAGGG 540

Qy  1514 ThrAlaSerSerGlyValSerIleAlaArgGlyValAlaProValIleValProGluLeu 1533    Qy 1091 ThrAlaArgProValLeuProArgProProThrIleSerAsnProProLeuLeuSer 1110
              ||||  :::||||||||||||||||||||||||||||||||||||||||||||||||            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACCCCAGCAAGCTCGGGGGGCTCCATTGCGCGCGGCGCCCGTCATTGTGCCTGAGCTG 600        Db   1 ACTGCCGCCGCCGTCCTGCCGCGGCCGCCCACCATCTCCAACCCGCCTCCCCTCATCTCC 60

Qy  1534 GlyLysProArgGlnSerProLeuThrTyrGluHisGlyAlaAspHisGlyAlaProPheAlaProAlaGlyHis 1553    Qy 1111 SerAlaLysHisProSerValLeuGluArgGlnIleGlyValAlaIleSerGlyAlaIleLeuGlyLysMetSer 1130
              ||||||||||||||||   ::::||||||||||||||||||||||||||||||||||                     |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGTAAGCCGCGGCAGAACCCCCTGACCTATGAGCACGGGGCACCCTTTGCCGGCCAC 660        Db  61 TCTGCCAAGCACCCGAGGGTCCTGGAGAGGCAAATAGGTGCCATCTCCCAAGGAATGTCG 120

Qy  1554 LeuProArgGlySerProValThrMetArgGluThrPro-ArgProLeuGlnGluGlyGly-- 1572    Qy 1131 ValGlnLeuHisValProTyrSerGluHisAlaAlaLysAlaProValGlyProValThrMet 1150
              ||||||||||||||||||||||||||||||||||   |||||||||||||||||||              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CTCCCACGAGGTTCGCCCGGTCACCCCCGGGAGCCCAGCCGCCGGCCTTGCAAGAGGGCA 720        Db 121 GTTCAGCTGCACGTGCCGTACTCAGAGCATGCAAGCATGCCCGGTGGCCCGTGTCACCATG 180

Qy  1573 ------SerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProAr 1590    Qy 1151 GlyLeuProLeuProMetCAspProLysLeuLysLeuAlaProPheSerGlyValValLysGlnGlu 1170
              :::    ||||||||||||||||||||||||||||       ||||||||                   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GGCTTTTCTCCCAGCGAAGGGCATCCCCAGGACCCCAGAAGCCTTGACGTCCAACCCCT 780        Db 181 GGCTGCCCTGCCCCATGACCCCAAAAAGCTGCACCCTTCAGCGAGTGAAGCAGGAGGAG 240

Qy  1590 gGluIleAlaLysSerProHisSerThrValProGluHisHisProHisProIleSerPr 1610    Qy 1171 GlnLeuSerProArgGlyAlaGlnAlaGlyThrAlaLeuGlySerValProGlyGlySerValProIleThrLys 1190
              :::   ||||  ||||| |||:::|||||||||||||||||||||||||||                     ||||||||||||||||||||||||||||||||||||||||||||||
Db   781 T---CTTGGAAAATGCCCAA-AGTCCCCGGCCGAGCAACCGGGGGGCCCGAAAAAACC 836        Db 241 CAGTTGTCCCACCGGAGGCAGGCTGGGCGTGGGGGTGCCCACAGCCCAG 300

Qy  1610 o 1610                                                                     Qy 1191 GluAlaSerValLeuArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLys 1210
              -                                                                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   837 A 837                                                                     Db 301 GAGGCGTCCGTGCTGAGAGGGACAGCTCTGGGCTCAGTTCCGGGCGAAGCATCACCAAA 360


RESULT 51                                                                         Qy 1211 GlyIleProSerThrArgValProSerAspSerAlaIleThrTyrArgGlySerIleGlySerIleThr 1230
BG252161                                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||
LOCUS     BG252161         876 bp   mRNA    linear   EST 13-FEB-2001           Db 361 GGCATTCCCAGCACCACGGGTGCCCTCGGACAGCGCCATCACCATACCGCGGCTCCATCACC 420
DEFINITION 602365028F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4473271 5',
          mRNA sequence.                                                        Qy 1231 HisGlyThrProAlaAspValLeuTyrLysTyrArgLeuTyrArgIleIleGluGlyValGluAsp 1250
ACCESSION BG252161                                                                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
VERSION   BG252161.1   GI:12761977                                             Db 421 CACGGCACGCCAGCTGACGTCCTGTACAAGGGACCACCATCACCAGGATCATCGGCGAGGAC 480
KEYWORDS  EST.
SOURCE    Homo sapiens (human)                                                 Qy 1251 SerProSerArgLeuAspArgArgGlyLeuAspSerLeuProLysGluAspSerLeuProLysGluThrTyr 1270
  ORGANISM Homo sapiens                                                               |||||||||||||||||||||||||||||||||||||||||||||||||||||
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   Db 481 AGCCCGAGTCGCTTGGACCTGGCCGCGGCCGGGAGGACAGCCTGCCCAAGGCCACGTCATCTAC 540
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)                                                    Qy 1271 GluGlyLysLysGlyHisGlyHisGlyValHisLeuSerTyrGluValThrGlnMetSerValThrGlnCysSer 1290
  AUTHORS NIH-MGC http://mgc.nci.nih.gov/.                                            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)       Db 541 GAAGGCAAGAAAGGGCCACGTCTTGTCCTATGAGGGTGGCATGTCTGTGACCCAGTGCTCC 600
  JOURNAL Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.                                     Qy 1291 LysGluAspGlyLysArgSerSerGlyProProHisGluThrAlaAlaProLysArgArgThr 1310
          Email: cgapbs-r@mail.nih.gov                                               ||||||||||||||||||||||||||||||||||||||||||||||||||||
          Tissue Procurement: ATCC                                             Db 601 AAGGAGGACGGCAGAAGCAGTCACGACCCCCATGAGACGGC-GCCCCCAAGCGCAC- 658
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)            Qy 1311 TyrAspMetMetGluGlyValGlyArgValAlaSerAlaSerIleGluGlyGlyLeu 1330
          DNA Sequencing by: Incyte Genomics, Inc.                                  |||||||||||||||||||||||||||||||||||||||||||||||||
          Clone distribution: MGC clone distribution information can be        Db 659 TATGACATGATGGAGGGCGTGGGCAGACCATCTCCTCAGC-AGCATTCGAAGGTCTC 717
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov                                                Qy 1331 Met-GlyArgAlaIleProProGluArgHisSerProHisHisLeuLysGluGlnHisHi 1350
          Plate: LLAM10295 row: h column: 08                                        ||||| ||||||||||||||||||||||||||:::::::
          High quality sequence stop: 711.                                      Db 718 ATGGGGCCCGTGCCATCCCCCGGAGCGCGAAA-AAGCCCCCACTACAAAGAAGAGGCGGACCCCA 776
FEATURES    Location/Qualifiers
     source 1..876
            /organism="Homo sapiens"
            /mol_type="mRNA"
```

us-09-522-753-5.rst

US-09-522-753-5 (1-2517) x CB248833 (1-763)

```
Qy   1472 AspValArgSerLeuIleGlySerProGlyArgThrPheProProValHisProLeuAsp 1491
          ||||||||||||||| :::||||||||||||||||     ::|:|||||||||
Db      3 GACGTGCGCTCCATCATCGGCAGCCCGGGCCGGACCCTTTCCTGCCTGCACCCGTCTGGAC   62

Qy   1492 ValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluSerLeuLysSerArg     1511
          :::|||||||||||||||||:::|||||||:||:|||:||||||||||||||||||
Db     63 ATAATGCTGCACGCCCGGCACTGGAGCGTGCTGCTATGGAAGAGAGTCTGAAGAGCCGG    122

Qy   1512 ProGlyThrAlaSerSerSerGlyGlySerIleAlaAlaArgGlyAlaProValIleValPro 1531
          |||||||||||||||||||||||| :|||||||||||:|||||||||||||||||||
Db    123 TCAGGGACCAGCAGTGTGCAGGGGGCTCCATCACACGTGGGGCTCCAGTGTGCTGCCT    182

Qy   1532 GluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 GAACTGGGCAAGCCACGGCAAAGCCCACTGACTTACGAAGACCACGGGCACCCTTCACC   242

Qy   1552 GlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlu 1571
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 AGTCACCTGCCACGTGGCTCCCCTGTGACCACGAGGAGCCCACGCCGCCTTCAGGAA    302

Qy   1572 GlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeuThrSerThrProArgGlu 1591
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 GGCAGCCTCCTATCCAGCAAGGCGTCCCAGGACAGGCGAAGCTGACATCTACACCCCGGGAG 362

Qy   1592 IleAlaLysSerProHisSerThrValProGluHisHisProIleSerProTyr     1611
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 ATCGCCAAGTCCCCACACAGCACTGTGCCCGAGCACCACCCCATCTCCCCCTAT    422

Qy   1612 GluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
          ||||||||||||| :::||||||||||||||||||||||||||||||||||||||
Db    423 GAGCACTTGCTCCGGGGCGTGACTGGTGTGGACCTGTACCGGTGGTCACATCCCATTGCCC 482

Qy   1632 PheAspProThrSerIleProArgGlyIleProLeuAsp---AlaAlaAlaAlaTyrTyr 1650
          |||||||||||||||||||||||||||||||||||     :::
Db    483 TTTGACCCCACCTCCATACCCCGAGGGATCCCTCTGGAAGCAGCAGCCGCCAGCCTACTAC 542

Qy   1651 LeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProHisProTyrLeuIle 1670
          |||||||||||||| :::|||||||||||||||||||||||||||||||||||||
Db    543 CTGCCCCGGCACTTGGCCCCCAGCCCCACCTACCCACACCTGTACCCACCTTACCTCATC  602

Qy   1671 ArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleGluAsnAspTyrIle 1690
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 CGCGGCTACCCTGACACGGCGGCCCTGGAGAACCGGCAGACCATCATCGAGAACGACTACATC 662

Qy   1691 ThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeu 1710
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 ACCTCGCAGCAGATGCACCACAACGTGCCTCCGCCATGGCCCAGCGGTGTGACATGCTG 722

Qy   1711 ArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsn 1723
          ||||||||||||||||||||||||||||||| ||||||
Db    723 AGGGGTCTGTCACCGCGAGAGTCCTCGCTGGNCCTCAAT 761
```

RESULT 53
AL884718/c

LOCUS   AL884718   845 bp   mRNA   linear   EST 16-SEP-2002
DEFINITION AL884718 XGC-egg Silurana tropicalis cDNA clone TEgg021o04 3', mRNA
            sequence.
ACCESSION   AL884718
VERSION   AL884718.1   GI:22925992
KEYWORDS   EST.
SOURCE   Silurana tropicalis (western clawed frog)
  ORGANISM Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1 (bases 1 to 845)
  AUTHORS   Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
  TITLE   Sanger Xenopus tropicalis EST project 2002
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Taylor R
            Sanger Institute

```
Qy   1350 sIleArgGlySerIleThrGlnGlyIleProArgSerTyrValGluAlaGlnGluAspTyr 1370
          ||||||      :::|||||||||||     |||||||||||||||     ::  ::
Db    777 TCCGCGG---GTCATCACCACAGGATCCCCG---GGCCACGTTGAGGCCAGGAGGAAC-- 828

Qy   1370 rLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrProProPro 1386
          |||      :::::  |||      ||||||||
Db    829 ----ACCCGGCGTCGGAAGCCAGCCCCAAAAGGAGGGGCCCCGCCCCA 873
```

RESULT 52
CB248833

LOCUS   CB248833   763 bp   mRNA   linear   EST 15-JUL-2003
DEFINITION UI-M-EX0-byj-k-12-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone
            IMAGE:5719043 5', mRNA sequence.
ACCESSION   CB248833
VERSION   CB248833.1   GI:28337632
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 763)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source   1..763
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5719043"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EX0"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site 2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA
                     tail, is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instututes of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
ORIGIN
Alignment Scores:
Pred. No.:   763
Score: 1.56e-41   Length: 763
                      1224.50   Matches: 230
Percent Similarity: 94.86%   Conservative: 10
Best Local Similarity: 90.91%   Mismatches: 12
Query Match: 9.27%   Indels: 1
DB: 14   Gaps: 1
```

```
Hinton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TEgg021oo4.q1kT7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES             Location/Qualifiers
     source          1..845
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TEgg021oo4"
                     /dev_stage="egg"
                     /lab_host="Escherichia coli XL1-blue"
                     /clone_lib="XGC-egg"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from egg.
                     EcoRI-NotI cut cDNA was then ligated into pCS107 with
                     EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Alignment Scores:
Pred. No.:           3.94e-41            Length:              845
Score:               1216.00            Matches:             230
Percent Similarity:  90.78%             Conservative:        26
Best Local Similarity: 81.56%           Mismatches:          20
Query Match:         9.20%              Indels:              6
DB:                                     Gaps:                2

US-09-522-753-5 (1-2517) x AL884718 (1-845)

Qy    195  LysLysGlnGlnLeuGluGluGluAlaAlaLysProProGluProGlyLysProVal  214
Db    844  AAAAAACAGCAACAGCTTGAAGAAGAAGCAGCAAAACCTTCCAGAGCCAGGAAGCCCGTC  785

Qy    215  SerProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluAsn  234
Db    784  TCCCCTCCACCTGTGGAACAGAAGCATCCAGTAGTGCAAATTATTTATGATGAAAAC  725

Qy    235  ArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyLeuGluLeu  254
Db    724  CGGAAAAAAGCAGAGGAAGCACACAGAATCTTAGAAGGTCTAGGTCCAAAGTAGAGCTG  665

Qy    255  ------------ProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsn  269
Db    664  TCATTTCTTTGCAGCCTCTTTACAACCAGCCCTCAGACACAAAGTTTATCATGAAAAC  605

Qy    270  IleLeuAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHis  289
Db    604  ATCAAGACAACAGGTGAGGAAAAAACTAATACTGTTCTTAAGAGAAGAAAATCAT  545

Qy    290  AlaAlaArgLysGlnTrpLysPheCysGlnArgTyrAspGlnLeuMetGluAlaLeu  309
Db    544  GCTAGAAAAACTAAGGAACAGAACATTTGTCAGCGCTATGATCAGCTGATGGAGGCTTGG  485

Qy    310  GluLysLysValGluArgIleGluGluAsnAsnProArgArgArgAlaLysSerLysVal  329
Db    484  GAGAAGAAAGTTGACAGGATAGAAAATAATCCTCGGAGAAAAGCAAGGAGCAAAACC  425

Qy    330  ArgGluTyrTyrGluLysGlnPheProGluLeuArgLysGlnGluLeuGlnLeuArg  349
Db    424  CGAGAATATTATGAGAAGCAATTTCCTGAAATAAGGAAGCAGGAGCAGCAGGAGCGT  365

Qy    350  MetGlnSerArgValGlyLysGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGlu  369
Db    364  TTCCAG---AGGGTGGGCAGAGAGCTGGGCTGTCAGCAACCATAGCCAGGAGTGAA  308

Qy    370  HisGluValSerIleIleAspGlyLysSerGluGluAsnLeuGluLysGlnMet  389
Db    307  CATGAAATATCTGAAATTATTGATGGTCTTTTCTGAACAGGAGAATAATGAAAAGCAGATG  248

Qy    390  ArgGlnLeuAlaValIleProProMetLeuTyrAspAlaAspGlnAlaGlnLysIleLysPhe  409
```

(second column)

```
Db    247  CGGCCACTTCAGTCATTCCACCCATGAGTCATTCCACCCATGAGTGTTGATGCGGAGCAAAGGAGGGTAAAATTC  188
Qy    410  IleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMet  429
Db    187  ATTAATATGAATGGTTGATGGAGGATCCCATGAAGGTGTACAACGACCCGCAGTTTATG  128

Qy    430  AsnMetTrpSerGluGlnGlyLysGluThrPheArgGluLysPheMetGlnHisProLys  449
Db    127  AATGTGTGGACTGACCACCAGAAAAGGAGATTTTTAAAGAAAAATTTGTACAACACCCTAAA  68

Qy    450  AsnPheGlyLeuIleLeuAlaSerPheLeuGluGlyLysArgLysThrValAlaGluValLeuTyr  469
Db    67   AACTTTGGACTGATAGCCTCGTATCTGGAAAGAAGAAGACTGTTTCAGACTGTGTTTTATAT  8

Qy    470  TyrTyr  471
Db    7    TATTAT  2

RESULT 54
BQ214358
LOCUS       BQ214358              900 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7594821 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6062786
            5', mRNA sequence.
ACCESSION   BQ214358
VERSION     BQ214358.1  GI:20395758
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 900)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13334  row: n  column: 03
            High quality sequence start: 8
            High quality sequence stop: 588.
FEATURES             Location/Qualifiers
     source          1..900
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6062786"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2 kb.    Library constructed by Life
                     Technologies."

ORIGIN

Alignment Scores:
Pred. No.:                               Length:              900
Score:               1.32e-40            Matches:             255
Percent Similarity:  1204.00             Conservative:        11
Best Local Similarity: 90.17%            Mismatches:          18
Query Match:         86.44%              Indels:              11
DB:                  9.11%               Gaps:                4
                     13

US-09-522-753-5 (1-2517) x BQ214358 (1-900)

Qy    234  AsnArgLysLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyLeuGlnValGlu  253
```

TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1949   row: o   column: 01
             High quality sequence start: 13
             High quality sequence stop: 527.
FEATURES   Location/Qualifiers
     source   1..1132
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454552"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:   Length: 1132
Score: 3.83e-40   Matches: 260
Percent Similarity: 1195.50   Conservative: 11
Best Local Similarity: 73.64%   Mismatches: 60
Query Match: 70.65%   Indels: 38
DB: 9.05%   Gaps: 8
                     12

US-09-522-753-5 (1-2517) x BM910968 (1-1132)

Qy   2151 HisProGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer 2170
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9 CACCCCAGCAGATCAGCGCACCCCTGCCAGCCCCCCTTCTACTCCTTCCTGGGCCCAGC 68

Qy   2171 CysProValLeuAspLeuArgArgProProSerAspHisTyrLeuProProAspHis 2190
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   69 TGCCCCGTCCTGGACCTCCGCCGCCCACCAGTGACCTCTACCTTCCCGCCCCCGACCAT 128

Qy   2191 GlyAlaProAlaArgGlySerProHisSerGluGluGlyGlyLysArgSerProGluProAsn 2210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   129 GGTGCCCCGGCCCCGTCCCCCGGCTCCCCCACAGCGAAGGGGGCAAGAGTCTCCAGACCCAAAC 188

Qy   2211 LysThrSerValLeuGlyGlyValGlyGluAspGlyIleGluProValSerProGlyGly 2230
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   189 AAGACGTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTCTCCCCACCGGAGGGC 248

Qy   2231 MetThrGluProGlyHisSerArgSerArgSerAlaValAlaTyrProLeuLeuTyrArgAspGlyGlu 2250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   249 ATGACGGAGCCAGGCACTCCCGAGTCGTGTGTACCCCTGCTGTACCGGGGATGGGGAA 308

Qy   2251 GlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla 2270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   309 CAGACGGAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCCCAGCCAGCC 368

Qy   2271 PheProSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnIleGlu 2290
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   369 TTCTTCAGCAAGCTCACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATC 428

Qy   2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluThrTyrAsnIleSerGlnProGlyGly 2310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   ::: |||||||||
     18 TCTGGAGAGAAGGCTGCACATCGGATTCTGGAAGGCCTGGGGCCCCAGCGTGGAG 77

Qy   254 LeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluLysIleAsn 273
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78 CTGCCGCTGTACAACCAGCCCTCCGACACCCGGCAGTATCATGAGAACATCAAAATAAAC 137

Qy   274 GlnAlaMetArgLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGln 293
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   138 CAGGCGGATGCGGAAGAGCTAATCTTGTACTTCAAGAGGAGGAATCACGCTCGGAAACAA 197

Qy   294 TrpLysGlnLysPheCysGlnLeuArgTyrArgAspTyrAspGlnLeuMetGluAlaLeuGluLeuLysLysVal 313
         |||::|||||||||||||||||||||||||||||||||||| |||||||||||
Db   198 TGGGAGCAGAAGTTCTGCCAGCTCCGCTACCGCGACTATGACCAGCTCATGGAGAAGGTG 257

Qy   314 GluArgIleGluAsnAsnProArgArgArgArgLeuLysGluSerLysValArgGluTyrTyr 333
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   258 GAGCGCATCGAGAACAACCCCCGGCGGCGGCGCCAAGGAGCAAGGTGCGCGAGTACTAC 317

Qy   334 GluLysGlnPheProGluIleArgLysGluLeuGluLeuGlnLeuGluArgMetGlnSerArg 353
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   318 GAGAAGCAGTTCCCTGAGATCCGCAAGGAGCTCGAGCTGCAGGAGCGCATGCAGAGCAGG 377

Qy   354 ValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaAlaArgSerGluHisGluValSer 373
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 GTGGGCCAGCGGGCCAGTGGGCTGTCCATGTCGGCCGCCGCCCGCAGCGAGCACGAGGTGTCA 437

Qy   374 GluIleIleAspGlyTyrLeuSerGluGluIleGluLysAsnLeuGluLysGlnMetArgGlnLeuAla 393
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 GAGATCATCGATGGCCTCTCAGAGCAGGAGAACCTGGAGAAGCAGATGCGCCAGCTCGCC 497

Qy   394 ValIleProProMetLeuTyrArgAspAlaAspGlnGlnAsnGlnArgIleLeuLysPheIleAsnMetAsn 413
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   498 GTGATCCCGCCCATGCTGTACGACGCTGACCAGCAGCCCATCAAGTTCATCAACATGAAC 557

Qy   414 GlyLeuMetAlaAspProMetLysValTyrLysValAspArgGlnValMetAsnMetTrpSer 433
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   558 GGGCTTATGGCGGACCCCATGAAGGTGTACAAAGACCGCCAGGTCATGAACATGTGGAGT 617

Qy   434 GluGlnLysLysGluGluThrPheArgGluGlnMetHisProLysAsnPhe-GlyLe 453
         ||||| |||||||||||||||||||||||||||||| |||||
Db   618 GAGCAGGAGAAGGAGAACCTTCCGGGAGAAGTTCATGCAGCATCCCAAGAACTTTGGGCCT 677

Qy   453 uIleAlaSerPheLeuGluArgLysThrVal-AlaGluCysVal-LeuTyrTyrTyrLeu 472
         |||||||||||||||||||||||||||::|||||||
Db   678 GATCGGCCATCATTCCTGGAGAGGAAAAACAGTGGGCTGAGTGCGTCCCTCCATTACTACCCT 737

Qy   473 ---ThrLysLysAsnGluAsnTyrLys-SerLeu---ValArgArgSerTyrArgArgAr 490
         |||||||||||||||||| :::||||||||||||||
Db   738 GACTTAGGAAAAATGAGAACTATAAGAACCCTGGTGAGACCGAAGGCTATCCGGCGCCG 797

Qy   490 gGlyLys---------SerGlnIleGlnArgLysGlnProGlnGlnGlnGlnGlnGlnGln-GlnGlnGlnGlnGlnG 507
         |||:: ||||||:::|:||||||||||||
Db   798 CGGGCAAGAACCCAGCAGCAAAACAAAAGGCCGGAGCAGCAGCCACCAGCAGCCAGGAG 857

Qy   507 lnGlnGlnProMet---ProArgSerSerGln 517
         ||||:::::
Db   858 CCAGCAGCAGCCCCATTGCCCCCCGGCAGCAACCAG 892

RESULT 55
BM910968
LOCUS   BM910968   1132 bp   mRNA   linear   EST 12-MAR-2002
DEFINITION AGENCOURT_6617188 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454552
            5', mRNA sequence.
ACCESSION   BM910968
VERSION   BM910968.1 GI:19361347
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1132)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.

```
FEATURES             Location/Qualifiers
source               1..893
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6141949"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Library constructed by Life
                     Technologies."

ORIGIN

Alignment Scores:
Pred. No.:                     1..893
Score:                 9.27e-40            Length:           893
Percent Similarity:    1183.50             Matches:          246
Best Local Similarity: 87.94%              Conservative:     2
Query Match:           87.23%              Mismatches:       19
DB:                    8.96%               Indels:           15
                       13                  Gaps:             4
```

```
US-09-522-753-5 (1-2517) x BU186180 (1-893)

Qy   841  GluGluGlnLysProProAlaAlaGluLeuLeuAlaValAlaAspThrGlyLysAlaGluGlu  860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6  GAGGAGCAGAAGCCCCCGCCGGCTGAACTGCTCGCAGTGGCAGACACAGGGAAAGGCCGAGGAG  65

Qy   861  ProValLysSerGluCysThrGluGluGluGlyProAlaLysGlyLysGlyLysGlyAspAla  880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    66  CCCGTCAAGAGCGAGTGCACGGAGGAGGAGGGCCGGGCCAAGGGCAAGGACGCG  125

Qy   881  GluAlaAlaGluAlaThrAlaGluGlyAlaLeuLysAlaGluLeuLysGluGlyGlySer  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   126  GAGGCCGCTGAGGCCACGGCCGAGGGCGCGCTCAAGGCAGAGAAGAAGGAGGGCGGGGAGC  185

Qy   901  GlyArgAlaThrAlaLysSerSerGlyAlaProGlnAsnAspSerSerAlaThr  920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   186  GGCAGGGCCACCACAGCCAAGAGCTCGGGCGCCCCCAGACAGAACGACTCCAGTGCCACC  245

Qy   921  CysSerAlaAspGluValAlaGluGlyTyrAspProArgAlaAsnAlaSerProGlnLysPro  940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   246  TGCAGTGCAGCAGATGAGGTGGCAGAGGGCTATGACCCGCGAGCAAGAACCGGCTGCGTGTCCCA  305

Qy   941  ArgProSerLeuLeuThrProThrGlyAspProArgAlaAsnAlaSerProGlnLysPro  960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   306  AGGCCCAGCCTCCTCACCCGACTGGCAACCCCGGGCCAATGCCCAGAAGCCA  365

Qy   961  LeuAspLeuLysGlnLeuLysGlnArgGlnAlaAlaAlaIleProProIleGlnValThrLys  980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   366  CTGGACCTGAAGCAGCTGAAGCAGCGGCAGGCTGCCATCCCCCCATCCAGGTCACCAAA  425

Qy   981  ValHisGluProProArgGluAspAlaAlaProThrLysProProAlaProProAlaProPro  1000
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   426  GTCCATGAGCCCCCGGAGGACGCAGCTCCCACCAAGCCAGCTCCCCAGCCCCACCG  485

Qy  1001  ProProGlnAsnLeuGlnProGluSerAspAlaProGlnProGlySerSerProArg  1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   486  CCACCGCAAAACCTGCAGCCGGAGAGCGACGCCCCTCAGCCAGGCAGCAGCCCCCGG  545

Qy  1021  GlyLysSerArgSerProAlaProProAlaAspLysGluGlu                        1033
          ||||||||||||||||||||||||||||||||||||||||||
Db   546  GGCAAGAGCAGGAGCCCGGCACCCCCGCCGACAAGGAGGAGAGCCTGTGTTCTTC  605

Qy  1034  --AlaPheAlaAlaGluAlaGlnLysLysLeuPro-GlyAspProProCysTrpThrSerGl  1052
            |||||||||||||||||||||    |||:|||||||||||||||||
Db   606  CCAGCCTTCGCAGCCGAGCCCAGAAACCTGCCTGGGGAACCCCCCTTGTGGACTTCCGG  665

Qy  1052  YLeuProPhe-ProValProProArgGluValIle-LysAlaSerProHisAlaPro---  1070
          |||||||| |||         |||::||||||||||||||
Db   666  CCTGCCCTTCCCCGGGGAGGTGATCCAGGGCTCCAGGGCTCTCCCCATGCCCCCGGGA  725
```

```
Db   429  AACAAGAAGCTGAACACCCACACCGGAATGAGGCCTGAATACAATATCAGCCAGCCTGGG  488

Qy  2311  ThrGluIlePheAsnMetGlyAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln  2330
          ||||||||||||||||||||||||||||||||||||||||||
Db   489  ACGGAGATCTTCAATATGCCCCCCATCACCGGAACAGGGCTTATGACCTATAGAAGCCAG  548

Qy  2331  AlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgGlyLysAlaLeu  2350
          |||||||||||:||||||:|||||:|||||||||||||||||||
Db   549  GCCGTGCAGGAACATGCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAGGGAAGGCCCTC  608

Qy  2351  Met-GlyLysTyrAspGln-TrpGluLeuSerProProLeuSerAlaLeuAsnAla-PheAsn  2369
          ||: |||||||||||  ||||:||||:||||||||||||:||   |||:|||||
Db   609  CTGGGGTAAATATGACCAGGTGGAAGAGTCCCCGCCGCTCAGCGCCCCATGCTTTTTACC  668

Qy  2370  ProLeuAsnAlaSerAla--SerLeuProAlaAlaMetProIleThr-AlaAla------  2386
          ||||||
Db   669  CCTCTGAAATGCCAGGGCCAGGGCGTGCTATGCCCCTTACCCGCTGGCTGACG  728

Qy  2387  -AspGlyArgSerAspHisThrLeuThrSerProGlyGly-----------------  2400
          ||||||    |||  ||||||
Db   729  GGACGGGAGTGGACCACATCACCTTCACCTT-CCACCAGGCGGGCGGCCGGGGGAAGGGG  787

Qy  2401  -----------GlyLysAlaLysSerValSerGlyValArgProSerSerArgLysAlaLy  2415
          |||| ||
Db   788  CCCAGGGTTCTTCTTGGGGAAAAACCCCCAGCCCCGAAAAAGCCCAGTTCCCCCGGC  847

Qy  2415  bSerProAlaProGlyLeuAlaSerGlyAspArg-----------ProProSerVa  2430
          |||:::||          ||:::||
Db   848  CCCTCCCGGCCCTGGGGATTCTGGGGGAACCGGCCCACCCCTTCCGGCCCCTAAGGGG  907

Qy  2430  lSerSerValHisSerGluGluGlyAspCysAsnArgArgProLeuThrAsn---ArgVa  2449
          |||||||      |||||||||
Db   908  CAACTCCGCAGGAGAAGACTGGCAACCCCCCCCAAACCCGTCCTTAACCAACCCGCGCG  967

Qy  2449  lTrpGluAspArgProSerSer------AlaGlySerThrProPheProTyrAsnPro--  2466
          ||:               :::
Db   968  GTGGGGAGGCACCGGCCCTCTCTCCCGCGGGGGCTCCACCCCCTTTCCCCCTGTC  1027

Qy  2467  ----------LeuIleMetArgLeuAlaGlyValMetAlaSer--  2478
          |||                  :::
Db  1028  CCCGCGACACGCTCGCGTTCCCGCACCCCGCTCTAATCCGCCCCCTGGT  1087

Qy  2479  ------ProProPro  2481
          ||||||||
Db  1088  GCGGACGCCCCCTTCCG  1103
```

```
RESULT 56
BU186180
LOCUS       BU186180              893 bp    mRNA    linear    EST 04-SEP-2002
DEFINITION  AGENCOURT_7861320 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6141949
            5', mRNA sequence.
ACCESSION   BU186180
VERSION     BU186180.1  GI:22700164
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 893)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13462 row: p column: 14
            High quality sequence stop: 540.
```

Mon Apr 19 08:15:09 2004   us-09-522-753-5.rst

```
Qy   1071 AspProSerAlaPheSer-TyrAlaProProGlyHisProLeu---ProLeuGlyLeuHi 1089
          |||  ||||||||  |||| ||||||||||||||||||     ||  |||||||||
Db    726 CCCCCTCAAGCTTCTCCCTACGCTCCCCTGGGTCCCCCAGCCCCCTGGGCCCTCC 785

Qy   1089 sAspThrAlaArgProValLeuPro---ArgProThrIleSerAsnProPro 1107
          |||||||||||||||||||||||||   ::||||  ||  ||
Db    786 ATGAACACTGGCCCCGGCCCCGGTCCTGGCGGGCCCAACCCACCATTCTCCAAACCCG 843
```

RESULT 57
BU257626
LOCUS   BU257626   790 bp   mRNA   linear   EST 26-NOV-2002
DEFINITION   603414849F1 CSEQCHN38 Gallus gallus cDNA clone CHEST337b13 5', mRNA
             sequence.
ACCESSION   BU257626
VERSION   BU257626.1   GI:25518285
KEYWORDS   EST.
SOURCE   Gallus gallus (chicken)
ORGANISM   Gallus gallus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
             Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 790)
AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
             Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE   A Comprehensive Collection of Chicken cDNAs
JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE   22335534
PUBMED   12445392
COMMENT   Contact: Simon Hubbard
             Department of Biomolecular Sciences
             University of Manchester Institute of Science and Technology
             (UMIST)
             PO Box 88, Manchester, M60 1QD, UK
             Tel: 0161612008930
             Fax: 0161360409
             Email: Simon.Hubbard@umist.ac.uk.
FEATURES   Location/Qualifiers
source   1..790
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST337b13"
                     /dev_stage="36"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHN38"
                     /note="Organ: limbs; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site_2: NotI; This normalized library was
                     constructed from 1 million independent clones. cDNA
                     synthesis was initiated using an oligo(dT) primer, using
                     methylated C in the first strand synthesis reaction.
                     Following this first strand reaction, double-stranded cDNA
                     was blunted, ligated to NotI adapters, digested with
                     EcoRI, size-selected, and cloned into the NotI and EcoRI
                     compatible sites of a custom modified MCS of the
                     pBluescript (KS+) vector. The library was normalized in 2
                     rounds using conditions adapted from Soares et al., PNAS
                     (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
                     (1996): 791, except that a significantly longer
                     reannealing hybridization was used."

ORIGIN

Alignment Scores:
Pred. No.:   8.5e-40   Length: 790
Score: 1183.00   Matches: 243
Percent Similarity: 89.89%   Conservative: 6
Best Local Similarity: 87.73%   Mismatches: 8
Query Match: 8.95%   Indels: 21
DB: 13   Gaps: 3

US-09-522-753-5 (1-2517) x BU257626 (1-790)

```
Qy   259 GlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLys 278
```

```
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2 CAGCCTTCAGACACCAGCAATACCATGAAAACATTAAAATAAACCAGGCAATGCGGAAG 61

Qy  279 LysLeuIleLeuTyrPheLysArgArgAsnHisAlaAlaArgLysGlnTrpLysGlnLysPhe 298
        |||||||||||||||::||||||||:::|||||||||||||||||||||
Db   62 AAGTTAATTTTGTACTTTAAGAGGAGAAATCATGCTGCAAGCAGTGGAACAGAAGTTC 121

Qy  299 CysGlnArgTyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAsn 318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  122 TGCCAGCGCTATGACCAGCTGATGGAGGCCTTGGAAAAGAAAGTGGAGCGCATTGAAAAT 181

Qy  319 AsnProArgArgArgArgAlaLysSerLysValArgGluTyrTyrLysGlnPhePro 338
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 AACCCCCGGCGCGTGCCAAGGAGAGCAAAGTGCGGGAATACTATGAAGAAGCAATTCCCT 241

Qy  339 GluIleArgLysGlnArgGluLeuGlnGluArgMetGln----------SerArgVal 354
        |||||||||||||||||||||||||||||||            |||||||
Db  242 GAGATCCGAAGCAGCAGGGAACTGCAAGAGCGCATGCAAGAGCGGTTAGAGGAGAAGCGGGTA 301

Qy  355 GlyGlnArgGlySer--GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSer 373
        |||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db  302 GGACAGAGGGGCAGCGAGGGCTTCCCATGTCCGCTGCGGCAGGAGCACGAGGTGTCG 361

Qy  374 GluIleIleAspGlyLeuSerGluGlnAsnLeuGlyLysGlnMetArgGlnLeuAla 393
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  362 GAGATCATCGATGGGCTTCAGAGCAAGAGAACCTGGAGAAGCAGACAGATGCGCCAGCTCGCC 421

Qy  394 ValIleProProMetLeuTyrAspAlaAspGlnArgIleLysPheIleAsnMetAsn 413
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 GTCATCCCACCATGCTCTATGATGCTGACCAGCAGCCCATTAAATTCATCAACATGAAC 481

Qy  414 GlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSer 433
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 GGCCTCATGGACGACCCCATGAAGGTCTACAAGGACCGGCAGGTGATGAACATGTGGAGC 541

Qy  434 GluGlnGluLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPheGluLeu 453
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  542 GAGCAGGAAGAGAAGGAGACCTTCCGGGAAAAGTTCATGCAGCACCCAAAGAACTTTGGCCTG 601

Qy  454 IleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeuTyrTyrLeuThr 473
        ||||:||||||||||||||||||||||:::|||||||||||||||||||
Db  602 ATTGCCTCCTTTCTGGACAGGAAGACGGTGGCAGACTCGTCCGTCCTGTACTACCTGACC 661

Qy  474 LysLysAsnGluAsnTyrLysSerLeuValArgSerTyrArgArgArg--GlyLysS 493
        |||||||||||||||||||||||||:::||||||||||||
Db  662 AAGAAGAATGAGAACTACAAGAACCTCGTGAGGAGAATTACCGGCGCGTGGGGAAAAA 721

Qy  493 erGlnGlnGluLysAspGluLysGluLysGluLysAspAspLysLysGlnLysGlnProMetP 513
        ||||||||||||||||||||||||||
Db  722 GCCAGCAGCAGCAA---------------------ATGC 742

Qy  513 roArgSerSerGlnGluGluLysAspGluLysGluLysGluLysGlu 528
        ||||||||||||::::      |||||||
Db  743 CCGG-AGCAGACAAGAGGAACAGACGAATGAACAGAAAGAAAGGAG 788
```

RESULT 58
BQ070408
LOCUS   BQ070408   1032 bp   mRNA   linear   EST 02-APR-2002
DEFINITION   AGENCOURT_6839381 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5922788
             5', mRNA sequence.
ACCESSION   BQ070408
VERSION   BQ070408.1   GI:19899454
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1032)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov

```
Db   589  GAGCTGGTGCCGCCACGGCTGTCCAAGAGGAGGAGCTGATCCAGAACATGACCGGCTGGAC  648
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181  ArgGluIleThrMetValGluGlnIleSerLysGluSerLysLeuLysLysLysGlnGlnLeu  200

Db   649  CGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGAACCAGCAACAAGCTG  708
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   201  GluGluGluAlaAlaLysProGluProProGluLysProValSerProProProIleGlu  220

Db   709  GAGGAGGAGGCTGCCAAGCCGCCAGCCTGAGAAACCCGTGTCACCGGCGCCATCGAG  768
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   221  SerLysHis---ArgSerLeuValGlnIleIleTyrAspGluAsn-ArgGlyLysAla-G  239

Db   769  TCGAAGCACCCCCAGCCTGGGTGCAAAATCTTCTCCACCGGAACACCCGGAAAAAGGCTTG  828
                |||||||   :::                   :||||                      |||||
Qy   239  luAlaAlaHisArgIleLeu--GluGlyLeuGly----ProGlnValGluLeuProLeu-  256

Db   829  AAGCTGGACATCGGAATTCCGGGAAAGGCTTGGGGGCCCAGGGGGAAACTCGCCGGCTG  888
                ||||||    :::::::                          :::||    |||||||
Qy   257  --TyrAsnGln--ProSerAspThrArg------GlnTyrHisGluAsnIleLysIleAsn  273

Db   889  GTACAACCAGACCCCTCCGAAAACCGGGGCGGTAATCATGAAGAACCTCTAAAATATAA  948
                |||||   |||||   |||    ||| |||:|                      :::
Qy   273  n 273

Db   949  C 949
```

RESULT 59
BX732682/c
LOCUS   BX732682   XGC-tadpole Silurana tropicalis cDNA clone TTpA071k07 3',
             858 bp   mRNA   linear   EST 18-NOV-2003
DEFINITION   BX732682 XGC-tadpole Silurana tropicalis cDNA clone TTpA071k07 3',
             mRNA sequence.
ACCESSION   BX732682
VERSION   BX732682.1 GI:3405423
KEYWORDS   EST.
SOURCE   Silurana tropicalis (western clawed frog)
  ORGANISM   Silurana tropicalis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
             Xenopodinae; Silurana.
REFERENCE   1 (bases 1 to 858)
  AUTHORS   Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
  TITLE   Sanger Xenopus tropicalis EST project 2001 (11_2003)
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Croning MDR
             Sanger Institute
             Hinxton, Cambridgeshire, CB10 1SA, UK
             Email: trop@sanger.ac.uk
             Sanger Xenopus tropicalis EST project 2001
             TROPICALIS_SEQUENCE_ID: TTpA071k07.q1kaT7
             Sequencing primer: T7
             This sequence is from a Xenopus Gene Collection (XGC) library
             constructed by Nigel Garrett.
             cDNA was oligo dT primed from 5ug of poly A+ RNA from tadpole
             embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
             EcoRI at the 5' end and NotI at the 3' end.
             Vector: pCS107; Site 1: EcoRI; Site_2: NotI
             Host: Escherichia coli DH10B.
FEATURES   Location/Qualifiers
     source   1..858
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TTpA071k07"
                     /dev_stage="tadpole (stage 35-40)"
                     /lab_host="E. coli DH10B"
                     /clone_lib="XGC-tadpole"
                     /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from tadpole
                     embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
                     with EcoRI at the 5' end and NotI at the 3' end"
ORIGIN

Alignment Scores:

Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2090 row: d   column: 21
High quality sequence stop: 623.

FEATURES   Location/Qualifiers
     source   1..1032
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5922788"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH MGC 47"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   2.44e-39   Length: 1032
Score: 1175.00   Matches: 246
Percent Similarity: 89.75%   Conservative: 8
Best Local Similarity: 86.93%   Mismatches: 19
Query Match: 8.89%   Indels: 10
DB: 13   Gaps: 5
```

US-09-522-753-5 (1-2517) x BQ070408 (1-1032)

```
Qy    1  MetSerGlySerThrGlnLeuValAlaAlaGlnThrTrpAlaThrGluProArgTyrPro  20
            |||||||||||  ||||||||     |||||||||||||||||||||||||||||||||
Db  109  ATGTCGGGGTCCACACAGCCTGTGGCACAGACGTGGAGGGCCACTGAGCCCCGCTACCCG  168

Qy   21  ProHisSerLeuSerTyrProValGlnIleAlaArgThrHisThrAspValGlyLeuLeu  40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  169  CCCCACAGCCTTTCCTACCCAGTGCAGATCGCCCGGACGCACACGGACGTCGGGCTCCTG  228

Qy   41  GluTyrGlnHisHisSerArgAspTyrAlaSerHisLeuSerProGlySerIleIleIleGln  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  229  GAGTACCAGCACCACTCCCGGGACTATGCCTCCCACCTGTCGCCCGGCTCCATCATCCAG  288

Qy   61  ProGlnArgArgProSerLeuLeuSerGluPheGlnProGlyAsnGluArgSerGln  80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  289  CCCCAGCGGCGGCCGAGCCCTCCCTGCTGTCTGAGTTCCAGCCCGGGAATGAACGGTCCCAG  348

Qy   81  GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet  100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  349  GAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCGAGCTGGGGAAGTCAGAGATG  408

Qy  101  GluPheIleGluSerLysArgProArgLeuLeuSerGluProLeuLeuLeuArgArgPro  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  409  GAGTTCATTGAAAGCAAGCGCCCTCGGCTAGAGCCTGCTGCCCCCCTGTTGCGACCCG  468

Qy  121  SerProLeuLeuAlaThrGlyIleGlnProAlaGlySerGluThrLysAspArgSer  140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  469  TCACCCCTGCTGGCCACGGGAATCCAGCCTGCGGGATCTGAAGACCTCACCAAGGACCGTAGC  528

Qy  141  LeuThrGlyLysLeuGluProValSerProProThrHisProHisProGluLeu  160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  529  CTGACGGGCAAGCTGGAACCGGTGTCTCCCCCCAGCCACACTGACCCCTGAGCTG  588

Qy  161  GluLeuValProProArgLeuSerLysGlyGluLeuIleGlnAsnMetAspArgValAsp  180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 60
BE793487
LOCUS       BE793487              787 bp    mRNA     linear   EST 20-SEP-2000
DEFINITION  601588814F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943059 5',
            mRNA sequence.
ACCESSION   BE793487
VERSION     BE793487.1  GI:10214789
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM797 row: 1 column: 04
            High quality sequence stop: 762.
FEATURES             Location/Qualifiers
     source          1..787
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3943059"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_7"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:                              Length:               787
Score:          8.75e-39                Matches:              246
Percent Similarity:  96.11%             Conservative:           1
Best Local Similarity:  95.72%          Mismatches:             8
Query Match:         8.77%              Indels:                 7
DB:                                     Gaps:                   1

US-09-522-753-5 (1-2517) x BE793487 (1-787)
```

```
Qy    2215  LeuGlyGlyGlyGluAspGlyIleGluProValSerProGluGlyMetThrGluPro 2234
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4  CTGGGTGGTGGTGAGGACGGTATTGAACCTGTCCCACCGGAGGGCGATGACGGAGCCA   63

Qy    2235  GlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspGlyGlnThrGluPro 2254
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64  GG-CACTCCCGAGTGTGTCTGTACCCGGATGGGAACAGACGGACGGAGCCC 122

Qy    2255  SerArgMetGlySerLysSerProGlyAsnThrSerGlnProAlaPhe-PheSerLy 2274
            |||||||||||||||||||||||||||||||||||||||||||||::::||
Db     123  AGCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCGCCAGCCTTCTACAGCAA 182

Qy    2274  sLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLysLysLysGlu 2294
            ||||||||||||||||||||||||||||||||||||||||||:::::
Db     183  GCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAAGAAGCT 242

Qy    2294  uAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrProGlyIlePh 2314
```

```
Pred. No.:
Score:              7.23e-39          Length:          858
Percent Similarity:  1161.50          Matches:         223
Best Local Similarity:  86.38%        Conservative:     37
Query Match:         74.09%           Mismatches:       24
DB:                  8.79%            Indels:           17
                     13               Gaps:              3

US-09-522-753-5 (1-2517) x BX732682 (1-858)
```

```
Qy     232  AspGluAsnArgLysAlaGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGln 251
            |||||||||||||||||       |||||||:||||||||||||||||:::
Db     858  GATGAAAACCGGAAAAAGACAGAGAAGCACCATAAGATCTTAGAAGGTCTAGGTCCAAAA 799

Qy     252  ValGluLeuProLeuTyrArgGlnProSerAspThrArgGlnTyrHisGluAsnIleLys 271
            |||||||||||||||||||||||||||||||||:::
Db     798  GTAGAGCTGCCTCTTTACAACCAGCCCTCAGACACAAACATCAAG 739

Qy     272  IleAsnGlnAlaMetArgLysLysLeuIleLeuTyrPheLysArgArgAsnHisAlaArg 291
            |||||||||||||||||||||||||||||||||||||||||||
Db     738  ACGAATCAGGTGATGAGGAAAAAACTAATACTGTTCTTTAAGAGAAGAAATCATGCTAGA 679

Qy     292  LysGlnTrpLysPheCysGlnArgTyrAspGlnMetGluMetGluAlaLeuGluLys 311
            |:::|||||||:||||||||||||||||||||||
Db     678  AAACTAAGGGACCAGAACATTTGTCAGCGCTATGATCAGCTGATGATGGAGGCTTGGGAGAAG 619

Qy     312  LysValGluArgIleGluAsnAsnProArgArgArgAlaLysSerGluLysValArgGlu 331
            ||:::::|||:::||||||||||||||||||||||||
Db     618  AAAGTTGACAGGATAGAAAAATAATCCTCGGAAAAAGGAGAGCAAAACCCGAGA 559

Qy     332  TyrTyrGluLysSerGlnPheProGluIleArgLysGlnArgGluLeuGlnLeuArgMetGln 351
            ||||||||||||||       ||||||||||||||
Db     558  TATTATGAGAAGCAATTTCCTGAAATAAGGAAGCAGAGAGAGCAGGAGCCGTTTCCAG 499

Qy     352  SerArgValGlyGlnArgGlySerMetSerArgAlaAlaArgSerGluHisGlu 371
            |||:::          |||||::  :::
Db     498  ---AGGGTTGGGCAGAGAGCTGGGCTGTCAGCAACCATAGCCAGGAGTGAACATGAA 442

Qy     372  ValSerGluIleIleAspProArgGluGlnGlyAsnLeuGluLysGlnLysMetArgGln 391
            :::||       ||||:||||||||||||||||||||
Db     441  ATATCTGAAATTATTGATGGTCTTTCTGAACACGAGAATAATCGAAAAGCAGATGCGCCAA 382

Qy     392  LeuAlaValIleProProMetLeuTyrArgAspAlaAspGlnArgIleLysPheIleAsn 411
            ||:::::||||||||||||:::|||:::||
Db     381  CTCTCAGTCATTCCACCCATGATGTTTGATGCGAGCAAAGAGGGGTAAAATTCATTAAT 322

Qy     412  MetAsnGlyLeuMetAlaAspProMetLysValTyrLysValAspArgGlnValMetAsnMet 431
            ||||||||||||       ||||:::
Db     321  ATGAATGTTTGATGGAGATCCCATGAAGGTGTACAAGGACCGGCCAGTTATGAATGTG 262

Qy     432  TrpSerGluGlnLysGluThrPheArgGluLysPheMetGlnHisProLysAsnPhe 451
            ||||||||||||       |||||||||||:::
Db     261  TGGACTGAGCAGAAAGAGACGTTTAAAGAGAAAATTTGTACAACACCCTAAAAACTTT 202

Qy     452  GlyLeuIleAlaSerPheLeuGluArgLysThrValAlaAlaGluLysValLeuTyrTyrTyr 471
            ||||||||||||       |||||:::||
Db     201  GGACTGATAGCCTCGTATCTGAAAGAAAGACTGTTTCAGACTGTGTTTTATTATTATTAT 142

Qy     472  LeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyr---ArgArgArg 490
            ||||||       ||||||:::||||
Db     141  TTAACTAAGAAAAATGAAAATTTTAAAGCCCTTGTGAGAAGAAACTACCCTAAAAGGAGA 82

Qy     491  GlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 510
            |||:::
Db      81  GGACGCCAACCAGCAACAG-------------- 64

Qy     511  ProMetProArgSerSerGlnGluGlnLysGluLysLysAspGluLysLysGluAlaGlu 530
            |:::       :::||||
Db      63  ---ATAACCCGCCCTGCCCAAGAAGAAGAAATGAAAAGTAGAAGAAAAAAAAAA 7

Qy     531  Lys 531
            |||
Db       6  AAA 4
```

/clone_lib="NIH_BMAP_FY0"
/note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGGGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   1.27e-38   Length: 718
Score: 1153.50   Matches: 216
Percent Similarity: 94.14%   Conservative: 9
Best Local Similarity: 90.38%   Mismatches: 13
Query Match: 8.73%   Indels: 1
DB: Gaps: 1

US-09-522-753-5 (1-2517) x CA749602 (1-718)

Qy 1480 ProGlyArgPhePheProProValHisProLeuAspAlaAspAlaArgAlaLeu 1499
          |||||||||||||||||||||||| ||||:||||||||||||::||||||||||||
Db   2 CCCGGCCGGCCTTTCCTCCGCCTGGACACCCGCTGGACATAATGGCTGACGGCCCGGGCACTG 61

Qy 1500 GluArgAlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerGly 1519
          ||||||||||||||||||||||||||:||:||||||||:|:|::|||:::||||||
Db   62 GAGCGTGCCTGCTATGAAGAGTCTGAAGAGCCGGTCAGGGACCAGCAGTGGTCAGGG 121

Qy 1520 GlySerIleAlaArgGlyAlaProValIleValAlaProGluLeuGlyLysProArgGlnSer 1539
          |||||||||||||||||||||||||||||||||:||:::||::|||||||||||||||||
Db   122 GGCTCCATCACAGTGGGGTCCAGGTCCCAGTTCGTCTGCTGACCAGCCAAGCCACGCGCAAAGC 181

Qy 1540 ProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerPro 1559
          |||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db   182 CCACTGACTTACGAAGACCACGGGGCACCCTTCACCAGTCACTGCCCACGTGGCTCTCCCT 241

Qy 1560 ValThrMetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAla 1579
          ||||||| |||||||:||:||||||||||||:||||||:|:|||||||||||||:
Db   242 GTGACCACGAGGGAGCCCACGCCCCGCTTCAGGAAGGGCAGCCTCCTATCCAGCAAGGCG 301

Qy 1580 SerGlnAspArgLysLeuThrSerHisArgGluIleAlaLysSerProHisSerThr 1599
          ||||||||||||||||||||||||||||||::||:||||||||||||||::||||||
Db   302 TCCCAGGACCGGAAGCTGACATCTCACCACCCGGGAGATGCCAAGTCCCCACACAGCACT 361

Qy 1600 ValProGluHisHisProIleSerProTyrGluHisLeuLeuArgGlyValSer 1619
          ||||||||::||:|||||||| ||||::||||::||||||||||||||||||:::
Db   362 GTGCCCGAGCACCACCCGATCTCCCCATATGAGCACTTGCTCCGGGGCGTGACT 421

Qy 1620 GlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerIleProArg 1639
          ||||||||:||||||||:|||||||||||||||||||:|||::||||||||||:|:||
Db   422 GGTGTGGACCTGTACCGTGTCACATCCATTGCCTTTGACCCCACCTCCATACCCCGA 481

Qy 1640 GlyIleProLeuAsp---AlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsn 1658
          ||||||||||::::: |||||||||||||||||||||||||||||||||||||
Db   482 GGGATCCCCTGGACGAAGCAGCCGCAGCCTACTACCTGCCCCGGCACTTGGCCCCCAGC 541

Qy 1659 ProThrTyrProHisLeuTyrProProTyrLeuIleArgGlyTyrProAspThrAlaAla 1678
          ||||||||||:||||||||||||||:|:|||:|||||||||||||||||||||||||||
Db   542 CCCACCTACCCACACCTGTACCCACCTTACCTCATCCGCGGCTACCCTGACACGGCGGCC 601

Qy 1679 LeuGluAsnArgGlnThrIleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsn 1698
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 CTGGAGAACCGCCAGACCATCATCAATGACTACATCACCTCGCAGATGCACCACAAC 661

Qy 1699 ThrAlaThrAlaMetAlaGlnArgGlnAspMetLeuArgGlyLeuSerProArgGlu 1717

Db   243 GAACACCCCACAACCGGAATGAGCCTGATACATATCAGCCAGCCTGGACGGAGATCTT 302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 2314 eAsnMetProAlaIleThrGlyTyrLeuMetThrArgSerGlnAlaValGlnGl 2334
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 CAATATGCCCGCCATCACCGGAACAGGCCTTATGACCTATAGAGAGCCAGGCGGTGCAGGA 362

Qy 2334 uHisAlaSerThrAsnMetGlyLeuGluAlaIleAlaIleLeuIleAlaLeuMetGlyLysTy 2354
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 ACATGCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGGTAAATA 422

Qy 2354 rAspGlnTrpGluGluSerProLeuSerGluAlaAsnAlaPheAsnProLeuAsnAlaSe 2374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423 TGACCAGTGGGAAGAGTCCCCGCTCAGCGCCAATGCTTTTAACCCTCTGAATGCCAG 482

Qy 2374 rAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerAspHisThrLe 2394
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483 TGCCAGCCTGCCCGCTGCTATGCCCATAACCGCTGCTGACGGACGGAGTGACCACACACT 542

Qy 2394 uThrSerProGlyGlyTyrGlyGlyTyrGlyLysAlaLysSerGlyValSerGlyArgProSerSerArgLysAl 2414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543 CACCTCGCCAGGTGGCGGCGGAAGGCCAAGGTCTCTGCAGACCCAGCCGAAAAGC 602

Qy 2414 aLysSerProAlaProGlyLeuAlaSerGlyAlaSerGlyAspArgProProSerValSerSerValHi 2434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   603 CAAGTCCCCCGGCCCCGGGGCCTGGCATCTGGGGACCGGCA-CCCTCGTCTCCTCAGTGCA 661

Qy 2434 sSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrpGluAspArgPr 2454
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   662 CTCGGAGGGAGACTGCAACCG--CGGACGGCCGTC-CCGACCACCGCGTGTGGGAGGACAGGCC 719

Qy 2454 oSerSerAlaGlySerThrPhePheProTyrAsnProLeuIleMetArg 2470
          ||||||||||||||||||| ||||||||||||||
Db   720 CTCGTCCCGCAG--TTCACCCCATTCCCT--AAATCCCTGATCATGCGG 763

RESULT 61
CA749602
LOCUS   CA749602   UI-M-FY0-cde-a-15-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
DEFINITION   IMAGE:6831736 5', mRNA sequence. 718 bp   mRNA   linear   EST 09-JUL-2003
ACCESSION   CA749602
VERSION   CA749602.1   GI:25571972
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 718)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
                National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE   Unpublished (1999)
JOURNAL
COMMENT   Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Dr. Jim Lin, University of Iowa
                cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                Clone Distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                This clone was contributed by the Brain Molecular Anatomy Project
                (BMAP)
                Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source   1..718
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6831736"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"

Db   662 GCTGCCTTCCGCCATGGCCAGGCGTGCTGACATGCTNAGGGGTCTGGTCACGCGAGAG 718
          |||::||||||||||||||||||||||||||||||||| ||||||||| |||||

RESULT 62
CD355812
LOCUS   CD355812   704 bp   mRNA   linear   EST 15-JUL-2003
DEFINITION   UI-M-FY0-cgo-o-10-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
             IMAGE:30355641 5', mRNA sequence.
ACCESSION   CD355812
VERSION   CD355812.1 GI:31148313
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 704)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
             National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
             http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)
             Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source   1..704
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30355641"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site 2: Not I; The library was constructed according
                     BonaIdo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   1.51e-38   Length: 704
Score: 1151.50   Matches: 218
Percent Similarity: 95.30%   Conservative: 5
Best Local Similarity: 93.16%   Mismatches: 10
Query Match: 8.71%   Indels: 1
DB: 14   Gaps: 1

us-09-522-753-5 (1-2517) x CD355812 (1-704)

Qy 1526 AlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyrGluAsp 1545
        |||||||||||||||||:|||||||:||||||||||||||||||||||||||||||||||
Db   3 GCTCCAGTCGTCGTGCCTGAACTGGGCAAGGCCACGCAAAGCCCACTGACTTACGAAGAC 62

Qy 1546 HisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArgGluPro 1565

Db   63 CACGGGGCACCCTTCACCGGTCACTGCCACGTGGCTCCCCTGTGACCACGAGGGAGCC 122
         |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
RESULT 62   1585

Qy 1566 ThrProArgLeuGlnGluGlySerLeuSerSerSerLysAlaSerGlnAspArgLysLeu 1585
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 123 ACGGCCACGCCTTCAGGAAGGCAGCGCCTCTATCCAGCAAGGCGTCCCAGGACCGGAAGCTG 182

Qy 1586 ThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHisHisPro 1605
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 183 ACATTCACACCCCGGGAGATCGCCAAGTCCCCACACACTGTGCCCGAGCACCACCCT 242

Qy 1606 HisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArg 1625
        |||||||||||||||||||||||||||||||||:|:|||||||||||||||||||||||
Db 243 CACCCCATCTCCCCATGAGCACTTGCTCCCGGGCGTGACTGGTGTGGACCTGTACCGT 302

Qy 1626 SerHisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAsp--- 1644
        ||||||||||||||||||:||||||||||||||:||||||||||||||||||||||:::
Db 303 GGTCACATCCCATTGGCCTTTGACCCCACCTCCATACCCCGGAGGGATCCCTCTGGAAGCA 362

Qy 1645 AlaAlaAlaAlaAlaTyrTyrLeuProArgHisLeuAlaProAsnProThrTyrProHisLeu 1664
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 363 GCAGCCGCAGCCTACTACCTGCCCCGGCACTTGGCCCCCAGCCCACTACCCACCACCTG 422

Qy 1665 TyrProProTyrLeuIleArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThr 1684
        ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||
Db 423 TACCCACCTTACCTCATCCGCGGCTACCCTGACACCGCGGCCCTGGAGAACCGGCAGACC 482

Qy 1685 IleIleAsnAspTyrIleThrSerGlnGlnMetHisHisAsnThrAlaThrAlaMetAla 1704
        |||||||||||||||||||||||||||:|||:|:||:||||||||||||||||||||||
Db 483 ATCATCAATGACTACATCACCTCGCAGCAGATGCACCACAACGCTGCTCCGCCATGGCC 542

Qy 1705 GlnArgAlaAspMetLeuArgGlyLeuSerProArgGluSerLeuAlaLeuAsnTyr 1724
        ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db 543 CAGCGTGCTGACATGCTGAGGGGTCTGTCACCCGAGAGTCCTTCGCCTCCCTCANTTAT 602

Qy 1725 AlaAlaGlyProArgGlyIleIleAspPLeuSerGlnValProHisLeuProValLeuVal 1744
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 603 GCCGCTGGNCCAAGAGGCATTATCGACCTGTCCCANGTGCCACACCTGCCCGTGCTGGTG 662

Qy 1745 ProProThrProGlyThrProAlaThrAlaMetAspArgLeu 1758
        |||||||||||||||||||||||||||||||||| ||||||
Db 663 CCACCAACGCCAGGCACCCCTGCCACCGCCATGGACAGGCTT 704

RESULT 63
BE728145
LOCUS   BE728145   804 bp   mRNA   linear   EST 15-SEP-2000
DEFINITION   601563413F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3833162 5',
             mRNA sequence.
ACCESSION   BE728145
VERSION   BE728145.1 GI:10142137
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 804)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
             National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
             Plate: LLCM511 row: i column: 03
             High quality sequence stop: 716.
FEATURES   Location/Qualifiers
     source   1..804
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

```
Qy  321 ArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPheProGluIle 340
            |||||| |||||:     :::::|  |||||
Db  717 CGGCGGGG-GCCAGGAGGCAGGGGGCGAGTAATA----------------------- 751

Qy  341 ArgLysGlnArgGluLeuGlnLeuArgMetGlnSerArgValGlyArgArgGlySerGly 360
            |||||||
Db  752 --------CGAGAGCAGGTCCTGAGAACGCAGCGGCGAGCTGGAGCGCGATGCGAGGGGG 802
```

RESULT 64
BM944466
LOCUS   BM944466   779 bp   mRNA   linear   EST 14-MAR-2002
DEFINITION UI-M-EH0p-bvr-k-08-0-UI.r1 NIH_BMAP_EH0p Mus musculus cDNA clone
            IMAGE:5695999 5', mRNA sequence.
ACCESSION   BM944466
VERSION   BM944466.1 GI:1928051
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 779)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source   1..779
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5695999"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 18.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_EH0p"
                     /note="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
                     Site 2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site. Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA
                     tail, is CAGCCACGAC. This library was created for the
                     University of Iowa Mouse Brain Developing Mouse Nervous
                     System; supported by National Institute of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
ORIGIN
```

Alignment Scores:
Pred. No.:   779
Score: 2.55e-37   Matches: 1123.00
Percent Similarity: 88.89%   Conservative: 7
Best Local Similarity: 86.21%   Mismatches: 26
Query Match: 8.50%   Indels: 5
DB: Gaps: 2

US-09-522-753-5 (1-2517) x BM944466 (1-779)

```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3831162"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_20"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN
```

Alignment Scores:
Pred. No.:   804
Score: 2.4e-37   Matches: 1124.00
Percent Similarity: 88.21%   Conservative: 4
Best Local Similarity: 86.79%   Mismatches: 22
Query Match: 8.51%   Indels: 16
DB: Gaps: 1

US-09-522-753-5 (1-2517) x BE728145 (1-804)

```
Qy   81 GluLeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMet 100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GAGCTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCGAGCTGGGGAAGTCAGAGATG 60

Qy  101 GluPheIleGluSerLysArgProArgLeuGluLeuGluProAspArgLeuLeuArgPro 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GAGTTCATTGAAAGCAAGCGCCCTCGGCTGGAGCTGCTGCCTGACCCCCTGCTGCGACCG 120

Qy  121 SerProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspArgLeuThrLysAspArgSer 140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TCACCCCTGCTGGCCACGGGCCAGCCTGCGGGATCTGAAGACCTCACCAAGGACCGTAGC 180

Qy  141 LeuThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeu 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CTGACGGGCAAGCTGGAACCGGTGTCTCCCCCCAGCCCCCCGCACACTGACCCTGAGCTG 240

Qy  161 GluLeuValProProArgLeuSerGluLysGluGlnIleGluLeuGlnAsnMetAspArgArgValAsp 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GAGCTGGTGCCGCCACGGCTGTCCAAGGAGGAGCTGATCCAGAACATGACCCGGTGGAC 300

Qy  181 ArgGluIleThrMetValGluGlnGlnIleLeuSerLysLeuLysLeuLysSerGlnGlnLeu 200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CGAGAGATCACCATGGTAGAGCAGCAGATCTTAAGCTGAAGAAGAAGCAGCAACAGCTG 360

Qy  201 GluGluGluAlaAlaLysProArgGluProGluLysProGluValSerProValIleGlu 220
            ||||||||||||||||||||||||||||  |||||||||||||||||||||||   |||
Db  361 GAGGAGGAGGCTGCCAAGCCGCGCGAG-CCTGAGAAGCCCGTGTCACCGGCCATC-GAG 418

Qy  221 SerLysHisArgSerLeuValGlnIleIleTyrArgGluTyrAspAsnArgLysLysGluAla 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  419 TCGAAGCACCGCAGCCTCGTTCAGATCATCTACCGACGAGCAGGAAGAGGCTGAAGCT 478

Qy  241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  479 GCACATCGGATTCTGGAAGGCCTGGGACCCCAGGTGGAGCTGCCGCTGTACAACCAGCCC 538

Qy  261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  539 TCCGACACCCGGCAGTATCATGAGAACATCAAAATAAACCAGGCGATGCGGAAGAAGCTA 598

Qy  281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCysGln 300
            |||||||||||||||||||||||||||||||||||||        |||  ::: ||||
Db  599 ATCTTGTACTTCAAGAGGAGGAATCACGCTCGGAAACAATGGGAGCAGAAGTC-TGCCAG 657

Qy  301 ArgTyrAspArgLeuMetGlnLeuLysLysGluGluValGlyArgIleGluGluAsnPro 320
            |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db  658 CGCTATGACCAGTC-ATGGAGGCTGGGAGGACGAAGGTGGAGAGCCATCGAGAGAACAACCCC 716
```

```
JOURNAL      Unpublished (2003)
COMMENT      Contact: Croning MDR
             Sanger Institute
             Hinxton, Cambridgeshire, CB10 1SA, UK
             Email: trop@sanger.ac.uk
             Sanger Xenopus tropicalis EST project 2001
             TROPICALIS_SEQUENCE ID: TEgg036f09.q1kT7
             Sequencing primer: T7
             This sequence is from a Xenopus Gene Collection (XGC) library
             constructed by Aaron M. Zorn.
             cDNA was oligo dT primed from 5ug of poly A+ RNA from egg.
             EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
             5' end and NotI at the 3' end.
             Vector: pCS107; Site_1: EcoRI; Site_2: NotI
             Host: Escherichia coli XL1-blue.

FEATURES             Location/Qualifiers
     source          1..811
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TEgg036f09"
                     /dev_stage="egg"
                     /lab_host="Escherichia coli XL1-blue"
                     /clone_lib="XGC-egg"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from egg.
                     EcoRI-NotI cut cDNA was then ligated into pCS107 with
                     EcoRI at the 5' end and NotI at the 3' end"

ORIGIN

Alignment Scores:
Pred. No.:                              Length:              811
Score:               5.2e-37            Matches:             216
Percent Similarity:  1116.00            Conservative:        25
Best Local Similarity: 88.93%           Mismatches:          23
Query Match:         79.70%             Indels:              7
DB:                  8.44%              Gaps:                2
                     13
```

US-09-522-753-5 (1-2517) x BX771375 (1-811)

```
Qy    2079 LysSerHisLeuGluGlyLeuArgProLysProGlnProGlyProValLysLeuGlyGly 2098
                :::      |||||||||||||||||||||   ::::||||
Db       2 GAAGAGCACTTGAAGGGGAGCTGCGGCCACAAGCAGCCAGGCCCATGAAGCTCAGCGCG 61

Qy    2099 GluAlaAlaHisLeuProHisLeuArgProLeuProGluSerGlnProSerSerSerPro 2118
           |||||||  ||| ||||  ||||||||||||||||||||||||||||||||||
Db      62 GAGGCTGCCCATCTCCCACATCTGCGGCCACTGCCCAGAGCCAGCCCTCATCTAGCCCA 121

Qy    2119 LeuLeuGlnThrAlaProGlyValValLysGlyHisGlnArgValValThrLeuAlaGlnHis 2138
           ||||||||||||||||||  ::::|||| ||||||||||||||||||||||||||||||
Db     122 CTCCTCCAGACTGCCCCAGGCATCAAAGGTCACCAGAGGGTGGTCACCTTGGCTCAGCAC 181

Qy    2139 IleSerGluValIleThrGlnHisAspTyrThrArgHisHisProGlnLeuSerAlaPro 2158
           ||||||||||||  ||||||||||||  |||||||||||||||||||||||||||
Db     182 ATCAGTGAGGTCATTACGCAGGACTACACGCGCCACCACCCCGAGCAGTCAGTGGCCCC 241

Qy    2159 LeuProAlaProLeuTyrSerArgPheProGlyAlaSerCysProValLeuAspLeuArgArg 2178
           ||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db     242 CTTCCCGCCCCTCTCACTCCTTTCCCGGAGCCAGCTGCCCTGTCCTGGATCTTCGCCGC 301

Qy    2179 ProProSerAspLeuTyrLeuProProAspHisGlyAlaAlaProAlaArgGlySerPro 2198
           ||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db     302 CCACCCAGTGACCTCTACCTCCCACCCCCCGACCATGGCACCCGCCCAGCCCGGGATCCCCC 361

Qy    2199 HisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGly 2218
           ||||||||||||||  ||||||||  ::: ||||||||||||||||||||
Db     362 CACAGTGAAGGGGGCAAAAGGTCCCCAGAACCCAAACAATCGGTCCTGGGCAGCAGC 421

Qy    2219 GluAspGlyIleGluProValSerProProGluGlyMetThrGluProGlyHisSerArg 2238
           |||||  ||||||||||||||  :::||
Db     422 GAGGATGCCCATTGAGCCTGGTGTCCCCAGCAGAGGGCATGACTGAGCCAGGACATGCTCGG 481

Qy    2239 SerAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGlyProSerArgMetGly 2258
           |||||||||||||||  ||||||||||||| |||||||
Db     482 AGCAGCTGGGTACCCACTGCTGTATCGAGACGGGGAACAGGGCCAGCCC---AGGATGGGC 538

Qy    2259 SerLysSerProGlyAsnThrSerGlnThrArgGlnProAlaPhe-PheSerLysLeuThrGluSe 2278
           ||||||||||||||  ||||||||||||||||||||
Db     539 TCTAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCCTTCCTTCAGTAGCTGACTGAGAG 598

Qy    2278 rAsnSerAlaMetValLysSerLysLysGlnGlnIleLeuGlnLysLysLeuAsnThrHisAs 2298
                 |    |||||||||||||||||||||||||||||||||||||||
Db     599 CAC-TCCGCCATGGTGAAGAGTCGAAGAAGCAGAGATCAACANGAAACTCAACACCCACAA 657

Qy    2298 nArgAsnGluProGlyTyrAsnIleSerGlnGlnProGlyTyrThrGluIlePheAsnMetProAl 2318
           |||||||||||||  |||||||||||||||||||||||||||||||||||||::
Db     658 CCGGAACAGCCAGAATACAATATTGGGCAGCCTGGGACGGAATCTTCANNCATGCGCC 717

Qy    2318 aIleThrGlyTyrGlyLeuMetThrTyrArgSerGlnAlaValAlaSerGlnHisAlaSerTh 2338
           |                  :::                   :::
Db     718 -ATCACTGGATACGGACAGGCTTATGANCCTGTAGAGCAGGCAGGGCGTGCAAGA---CACCGCCAGCAC 773

Qy    2338 r 2338
           |
Db     774 A 774
```

```
RESULT 65
BX771375/c
LOCUS       BX771375    811 bp    mRNA    linear    EST 10-DEC-2003
DEFINITION  BX771375 XGC-egg Silurana tropicalis cDNA clone TEgg036f09 3', mRNA
            sequence.
ACCESSION   BX771375
VERSION     BX771375.1    GI:39678580
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
ORGANISM    Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 811)
AUTHORS     Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
TITLE       Sanger Xenopus tropicalis EST project 2001 (11_2003)
```

```
Db    333  AGCTGGGCTGTCAGCAACCATAGCCAGGAGTGAACATGAAATATCTGAAATATTATTGATGG  274
           |:::||||||          ::      ||||||||||||||||||||:::|||||||||||

Qy    378  yLeuSerGluGlnLeuGluAsnLeuLysGlnMetArgGlnLeuAlaValIleProProMe   398
           ||||||||||||||||||||:||||||||||||||||||||:||||||||||||||||||
Db    273  TCTTTCTGAACAGGAGAATAATGAAAAGCAGATGCGCCAACTTTCAGTCATTCCACCCAT   214

Qy    398  tLeuTyrArgAspAlaAspGlnGlnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAs   418
           |:::::||||:::||:||||||||||||||||||||||||||||:||:||||||||||
Db    213  GATGTTTGATGCTGACAAGGAGAGGGTAAATTCATTAATATGAATGGTTTGATGGCAGGA   154

Qy    418  pProMetLysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnLeuLysGl   438
           |||||||||||||||||||||||||:::|||||:||||:      |||||||
Db    153  TCCCATGAGGTGTACAAGGACCCGGCAGTTTATGAATGTGTGGGACACACGAAAAGGA    94

Qy    438  uThrPheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLe   458
           |||||:::||||||||||||:::|||||||||||||||:||||||||:::|
Db     93  GATTTTTAAAGAAAAATTTGTACACACCCTAAAAACTTTGGACTGATAGCCTCGTATCT    34

Qy    458  uGluArgLysThrValAlaGluCysValLeu   468
           ||||||||||||||:|||||:::||||||
Db     33  GGAAAGAAAGACTGTTTCAGACTGTGTTTA    3
```

```
RESULT 66
CB521565
LOCUS        CB521565           774 bp    mRNA      linear    EST 09-JUL-2003
DEFINITION   UI-M-GH0-cen-c-02-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
             IMAGE:6841755 5', mRNA sequence.
ACCESSION    CB521565
VERSION      CB521565.1  GI:2935420
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 774)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/,
             National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE      Unpublished (1999)
  JOURNAL
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Dr. Jim Lin, University of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
             http://genome.uiowa.edu/distribution/mousefl.html
             This clone was contributed by the Brain Molecular Anatomy Project
             (BMAP)
FEATURES             Location/Qualifiers
     source          1..774
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6841755"
                     /tissue_type="Whole brain"
                     /dev_stage="1, 5, and 15 days newborn"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GH0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is CGAACTGAAT. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
```

ORIGIN

Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

```
Alignment Scores:
Pred. No.:             774
Score:                 1114.00
Percent Similarity:    92.69%
Best Local Similarity: 86.54%
Query Match:           8.43%
DB:

Length:                774
Matches:               225
Conservative:          16
Mismatches:            14
Indels:                8
Gaps:                  3
```

US-09-522-753-5 (1-2517) x CB521565 (1-774)

```
Qy   1244  ThrArgIleIleGluAspSerProSerArgLeuAspArgGlyArgGluAspSerLeu  1263
           :::||||||||||||||||||||||||||||||||||||:||||||::|||||::
Db      2  AGCAGGATCGTCGGTGAGGACAGCCCAAGTCGCCTTGACCGGCACGAGAGGACACCCTG    61

Qy   1264  ProLysGlyHisValIleTyrGluGlyLysGlyLysSerGlyHisValLeuSerTyrGluGlyLysGly  1283
           |||||||||||||||||:||::|||:|||||||||||||||||       :::::::
Db     62  CCCAAGGGCCATGTCATCTATGAGGGCAAGAAAGGCCACGTCCTATCCTATGAAGGTGGT   121

Qy   1284  MetSerValThrGlnCysSerLysGlyAspLysArgSerSerGlyGlyGlyGluProProHis  1303
           ||:::|||||||||||||||||||||||||:||||||||||||||||:::||||||||
Db    122  ATGTCCGTGTCACAGTGCTCTAAGGAGGATGGAAGGAGCAGCTCGGGCCCACCCCATGAG   181

Qy   1304  ThrAlaAlaProLysGlyTyrArgThrTyrAspMetMetGluGlyArgValGlyArgAlaIleSer  1323
           |||||||||||||||||||||:||||||||||||||:||||||:||:::::::
Db    182  ACTGCGCCCCTAAACGCCATGACCATGATGGAGGGCCGTGTAGGCAGGACTGTCACC   241

Qy   1324  SerAlaSerIleGluGlyLeuMetGlyArgAlaIleProProGluArgHisSerProHis  1343
           ||:::||||||||||||:||||||||||:::||:::|||||:|||||:::||
Db    242  TCAGCCAGCATAGAGGGACTCATGGGCCGCGCCCATC---CCTGAGCGACAGCAGCCCC--   295

Qy   1344  HisLeuLysGluGlnHisHisIleHisArgGlySerIleThrGlnGlyIleProArgSerTyr  1363
           |||||||||||:::||||:||||||:||||||||||||||:||::::::::::
Db    296  CACCTCAAGGAGCAGCATCACATCCGAGGCTCCATCACGCAAGGGATCCCGAGGTCCTAT   355

Qy   1364  ValGluAlaGlnGluAspTyrLeuArgArgGluAlaLysLeuLeuLysArgGluGlyThrGlyThr  1383
           ||||||||||||||||||||||||||:||||||:||||||||:||||||||||||||
Db    356  GTTGAGGCCCAGGAGGACTACTTACGGCGGGAGGCCAAGCTCTTGAAGCGAGAAGGGACA   415

Qy   1384  ProProProProSerArgAspLeuThrGluAlaTyrLysThrGln-----------  1400
           |||||||||||||||||||||||||||||||||:::
Db    416  CCACCACCCCCCACCACCACCTACGGGACCTGACACCTACAAGCCCCGCCCCCTGGAC   475

Qy   1401  AlaLeuGlyProLeuLeuProAlaHisGluGlyLeuValAlaThrValLysGlu  1420
           ||||||||||||||||||||:||||||:||||||||||||:::|||||::|||||
Db    476  CCCTCTGGGTCCCCTGAAGCTGAAGCCGACTCACGAGGGTGTGGTAGCAACTGTGAAGGAG   535

Qy   1421  AlaGlyArgSerIleHisGluIleProArgGluGlnLeuArgHisThrProGluLeuPro  1440
           |||||||||||||:::||||||||||||||||||||||:::||||||||||:||:
Db    536  GCGGGCCGCTCTATCCATGAGATCCCGAGAGAGCAGCTGCGGCACACCTGAGCTACCC   595

Qy   1441  LeuAlaProArgProLeuLysGluGlyLeuSerIleThrGlnGlyThrProLeuLysTyrAsp  1460
           ||||||||:||||||||||:::||||||||||||||||||||||||||||:||||||
Db    596  CTGGCACCACGGCCGCTTCCTGAAGGAGGGTTCCATCACCCAGGGCACCCCACTCAAGTACGAC   655

Qy   1461  ThrGlyAlaSerThrThrGlySerLysHisAspValArgSerLeuIleGlySerPro  1480
           :::|||||||||||:::         |||||||||:::|||:||||||||||:::||
Db    656  TCTGGGGCACC-TCCACTGGCCACCAN-GAACACGACGTGCGCTCATCATCGGCAG-CCC   712

Qy   1481  GlyArgThrPhePheProProValHisProLeuAspValMetAlaAspAlaArgLeuGlu  1500
           ||||| ||||  |||:| ::  ||||||||||||||||||||||||||||||||||
Db    713  GGNCGGCCCTTCCCTGCCGCCANCCGCTGGACATAATGGCTGACGCCCGGGCACTGGAG   772
```

```
RESULT 67
BM783748
LOCUS        BM783748           651 bp    mRNA      linear    EST 05-MAR-2002
DEFINITION   K-EST0061743 S5SNU484 Homo sapiens cDNA clone S5SNU484-36-A06 5',
             mRNA sequence.
ACCESSION    BM783748
```

```
VERSION     BM783748.1  GI:19131980
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002).
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 36  row: A  column: 06
            High quality sequence stop: 651.
FEATURES             Location/Qualifiers
     source          1..651
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SSSNU484-36-A06"
                     /sex="M"
                     /tissue_type="Stomach"
                     /cell_type="Epithelial"
                     /cell_line="SNU-484"
                     /lab_host="Top10F'"
                     /clone_lib="SSSNU484"
                     /note="Organ: Stomach; Vector: pTZ18RP1; Site 1: EcoRI;
                     Site 2: NotI; The poly (A)+ RNA was decapped with tabacco
                     acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
                     including EcoRI site by treatment of T4 RNA ligase. The
                     first strand cDNA was synthesized from oligo dT-selected
                     mRNA by priming with dT-tailed vector. The dT-tailed
                     vector was adjusted to have about 60nt. The cDNA vector
                     was circularized with E. coli DNA ligase after digestion
                     of EcoRI which site is also included in vector. An RNA
                     strand converted to a DNA strand by Okayama-Berg method.
                     The obtained cDNA vectors were used for transformation of
                     competent cells E. coli Top10F' by electroporation
                     method."

ORIGIN

Alignment Scores:
Pred. No.:
Score:                 6.57e-37       Length:          651
Percent Similarity:    1111.00        Matches:         215
Best Local Similarity: 81.75%         Conservative:    0
Query Match:           81.75%         Mismatches:      2
DB:                    8.41%          Indels:          46
                       12             Gaps:            1

US-09-522-753-5 (1-2517) x BM783748 (1-651)

Qy  2145 GlnAspTyrThrArgHisHisProGlnLeuSerAlaProLeuProAlaProLeuTyr 2164
         |||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAGGACTACACCCGGCACCACCCCAGCAGCTCAGCCGCACCCCTGCCCGCCCCCCTCTAC 60

Qy  2165 SerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAspLeuTyr 2184
         ||||||||||||||||||||||||||||||||
Db    61 TCCTTCCCTGGGGCCAGCTGCCCCGTCCTGGACCTCCGCCGGCCACCCAGTGACCTCTAC 120

Qy  2185 LeuProProAspHisGlyAlaArgGlySerProHisSerGluGlyGlyLys 2204
         ||||||||||||||||||||||||||||||||||||||
Db   121 CTCCCGCCCCGGACCATGGGCGCCCGCGGGTCGCTCCCCCACAGCGAAGGGGGCCAAG 180

Qy  2205 ArgSerProGluProAsnLysThrSerValLeuGlyValGlyAspGlyIleGluProPro 2224
         ||||||||||||||||||||||||||||||||||||||
Db   181 AGGTCTCCAGAGCCCAAACAGACCAGACCGTCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCT 240

Qy  2225 ValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaValValTyrProLeu 2244
         |||||||||||||||||||||||||||||
Db   241 GTGTCCCCACCGGAGGGCATGACGGACGGAGCCAGGGCACTCCCGGAGTGCTGTGTACCGCTG 300

Qy  2245 LeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsn 2264
         |||||||||||||||||||||||||||||||
Db   301 CTGTACCGGGATGGGAACAGACGAGCCCAGCAGGATGGGTCCAAGTCTCCAGGCAAC 360

Qy  2265 ThrSerGlnProProAlaPhePheSerLysSerLeuThrGluSerAsnSerAlaMetValLys 2284
         ||||||||||||||||||||||||||||||
Db   361 ACCAGCCAGCCGCCAGCCTTCTTCAGCAAGTCGACCGAGAGCAACTCCGCCATGGTCAAG 420

Qy  2285 SerLysGlyGlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyr 2304
         ||||||||||||||||||||||||||||||||
Db   421 TCCAAGAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATAC 480

Qy  2305 AsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeu 2324
         ||||||||||||||||||||||||||||||||
Db   481 AATATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACCGGAACAGGCCTT 540

Qy  2325 MetThrTyrArgSerGlnAlaValGlnHisGluHisAlaSerThrAsnMetGlyLeuGluAla 2344
         |||||||||||||||||||||||||||||||
Db   541 ATGACCTATAGAAGCCAGGCGGTGCAGGAACATGCCAGCACCAACATGGGCCTGGAGGGC 600

Qy  2345 IleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGlnSerProProLeuSer 2364
         ----------------
Db   601 CATATTAGAAAGGCCACTCATG--------------- 621

Qy  2365 AlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThr 2384
         ------------
Db   621 ------- 621

Qy  2385 AlaAlaAspGlyArgSerProAspHisThrSerProGlyGlyGlyGlyLysAlaLys 2404
         ||||||||||||
Db   622 ------GGTGCGGCGGCGGAAGGCCAAG 642

Qy  2405 ValSerGly 2407
         |||||||||
Db   643 GTCTCTGGC 651

RESULT 68
CF540568
LOCUS       CF540568                 683 bp    mRNA    linear   EST 12-SEP-2003
DEFINITION  UI-M-GV0-cie-18-0-UI.r1 NIH_BMAP_GV0 Mus musculus cDNA clone
            IMAGE:30545297 5', mRNA sequence.
ACCESSION   CF540568
VERSION     CF540568.1  GI:34593091
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE     Unpublished (1999)
  JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT     Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..683
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
```

Qy   1731 IleIleAspLeuSerGlnValProHisLeuProThrProGlyThr 1750
            ||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATTATCGACCTGTCCCAGTGCCACACCTGCCCGTGCTGGTGCCCACAACGCCAGGGCAC 660

Qy   1751 ProAlaThrAlaMetAspArg 1757
            |||||||||||||||||::
Db   661 CCTGCCACGGCCATCGACCGC 681

RESULT 69
AI830862
LOCUS   AI830862   622 bp   mRNA   linear   EST   21-DEC-1999
DEFINITION   wj61b09.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2407289 3'
             similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
             ; mRNA sequence.
ACCESSION   AI830862
VERSION   AI830862.1 GI:5451533
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 622)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Christopher Moskaluk, M.D., Ph.D.
             Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 2006   Std Error: 0.00
             Seq primer: -40UP from Gibco
             High quality sequence stop: 465.
FEATURES   Location/Qualifiers
     source 1..622
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:2407289"
             /tissue_type="squamous cell carcinoma, poorly
             differentiated (4 pooled tumors, including primary and
             metastatic)"
             /dev_stage="adult"
             /lab_host="DH10B (phage-resistant)"
             /clone_lib="NCI_CGAP_Lu19"
             /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
             modified polylinker; 1st strand cDNA was prepared from
             pooled lung tumor tissue, and was then primed with a Not I
             - oligo(dT) primer. Double-stranded cDNA was ligated to
             Eco RI adaptors (Pharmacia), digested with Not I and
             cloned into the Not I and Eco RI sites of the modified
             pT7T3 vector. Library went throughb one round of
             normalization. Library constructed by Bento Soares and M.
             Fatima Bonaldo. "

ORIGIN

Alignment Scores:
Pred. No.:   9.15e-37   Length: 622
Score: 1107.00   Matches: 207
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 8.38%   Indels: 0
DB: 9   Gaps: 0

US-09-522-753-5 (1-2517) x AI830862 (1-622)

Qy   2117 SerProLeuGlnThrAlaProGlyValLysGlyHisGlnArgValValThrLeuAla 2136

             /db_xref="taxon:10090"
             /clone="IMAGE:3054527"
             /tissue type="whole brain"
             /dev_stage="1,5, and 15 days newborn"
             /lab_host="DH10B (T1 phage resistant)"
             /clone_lib="NIH_BMAP_GV0"
             /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
             Site_2: Not I; The library was constructed according
             Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
             1996. Denatured RNA was size fractionated on a 1% agarose
             gel.First strand cDNA synthesis was primed with oligo-dT
             primer containing a Not I site. Double strand cDNA was
             size selected according to mRNA size fraction,ligated with
             EcoR I adaptor, digested with NotI and then cloned
             directionally into pYX-Asc vector. The library tag
             sequence located between the Not I site and the polyA tail
             is CGAACTGAAT. This library was created for the University
             Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
             Developing Mouse Nervous System', supported by National
             Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
             program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   7.27e-37   Length: 683
Score: 1110.50   Matches: 210
Percent Similarity: 94.71%   Conservative: 5
Best Local Similarity: 92.51%   Mismatches: 11
Query Match: 8.40%   Indels: 1
DB: 14   Gaps: 1

US-09-522-753-5 (1-2517) x CF540568 (1-683)

Qy   1532 GluLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAla 1551
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GAACTGGGCAAGCCACGGCAAAGCCCACTGACTTACGAAGACCACGGGGCACCCTTCACC 60

Qy   1552 GlyHisLeuProArgGlySerProValThrMetArgGluProThrProArgLeuGlnGlu 1571
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGTCACCTGCCACGTGGCTCCCCTGTGACCACGAGGGAGCCCACGCCCTTCAGGAA 120

Qy   1572 GlySerLeuSerSerLysAlaSerGlnAspArgLysThrSerThrProArgArgGlu 1591
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGCAGCCTCCTATCCAGCAAGGGGTCCCAGGACCGGAAGCTGACATCTTACACCCCGGGAG 180

Qy   1592 IleAlaLysSerProHisSerProThrValProGluHisHisProHisSerProProTyr 1611
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATCGCCAAGTCCCCACACAGCACTGTGCCGGAGCACCACCCTCACCCCATCTCCCCCTAT 240

Qy   1612 GluHisLeuLeuArgGlyValSerGlyValAspLeuTyrArgSerHisIleProLeuAla 1631
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GAGCACTTGCTCCGGGGCGTGACTGGTGTGGACCTGTACCGGTCGTCACATCCCATTGGCC 300

Qy   1632 PheAspProThrSerIleProArgGlyIleProLeuAsp--AlaAlaAlaAlaTyrTyr 1650
            ||||||||||||||||||::|||||||::
Db   301 TTTGACCCCACCTCCATACCCGGAGGATCCCCTCTGGAAGCAGCAGCCGCAGCCTACTAC 360

Qy   1651 LeuProArgHisLeuAlaProAsnProThrTyrProHisLeuTyrProProTyrLeuIle 1670
            ||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTGCCCCGGCACTTGGCCCCCAGCCCCACCTACCCCACCCTGTACCCACCTTACCTCATC 420

Qy   1671 ArgGlyTyrProAspThrAlaAlaLeuGluAsnArgGlnThrIleIleAsnAspTyrIle 1690
            ||||||||||||||||||||||||||||||||||||||||||||
Db   421 CGCGGCTACCCCTGACACCGGCGCCCTGGAGAACCGCCAGACCATCATCAATGACTACATC 480

Qy   1691 ThrSerGlnMetHisHisAsnThrAlaThrAlaMetAlaGlnArgAlaAspMetLeu 1710
            |||||||||||||::||||||||||||||||||||||||
Db   481 ACCTCGCAGATGCATGACCACAACGCTGCCTCCGCATGGCCCAGCGTGCTGACATGCTG 540

Qy   1711 ArgGlyLeuSerProArgGluSerSerLeuAlaLeuAsnTyrAlaAlaGlyProArgGly 1730
            |||||||||||||||||||||||||||||||::||||
Db   541 AGGGGTCTGTCACCCGCGGAGAGTTCCTCGCTGGNCCTCAATTATGCCGCTNGGCCANGAGGGC 600

```
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4214551"
                    /tissue_type="anaplastic oligodendroglioma with 1p/19q
                    loss"
                    /lab_host="DH10B (T1 phage-resistant)"
                    /clone_lib="NCI_CGAP_Brn67"
                    /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectional; Primer: Oligo dT.
                    Average insert size 2.3 kb. Constructed by Life
                    Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Alignment Scores:
Pred. No.:                  1.18e-36        Length:                 657
Score:                      1105.00         Matches:                210
Percent Similarity:         99.53%          Conservative:           0
Best Local Similarity:      99.53%          Mismatches:             1
Query Match:                8.36%           Indels:                 0
DB:                                         Gaps:                   0

US-09-522-753-5 (1-2517) x BF530324 (1-657)

Qy   1447 LysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAlaSerThrThr 1466
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    657 AAGGAGGGCTCCATCACGCAGGGCACCCCGCTCAAGTACGACACCGGCGCGTCCACCACT  598

Qy   1467 GlySerLysHisAspValArgSerLeuIleGlySerProGlyArgThrPheProPro 1486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 GGCTCCAAAAGCACGACGTACGCTCCCTCATCGGCAGCCCGGCGGTTCCACCC  538

Qy   1487 ValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCysTyrGluGlu 1506
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    537 GTGCACCCGCTGGATGTGATGGCCGACGCCCGGGCACTGGAACGTGCCTACGAGGAG  478

Qy   1507 SerLeuLysSerArgProGlyThrAlaSerSerGlySerIleAlaArgGlyAla 1526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 AGCCTGAAGAGCCGGCCAGGGACCGCCAGCAGTCGGGGGTCCATTGCGCGGGCGCC  418

Qy   1527 ProValIleValProLeuGlyLeuGlyLysProArgGlnSerProLeuThrTyrGluAspHis 1546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417 CCGGTCATTGTGCCTGAGCTGGGCAAGCCGCGGCAGAGCCCCCTAACCTATGAGGACCAC  358

Qy   1547 GlyAlaProPheAlaGlyHisLeuProLeuProArgGlySerProValThrMetArgGluProThr 1566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 GGGGCACCCTTTGCCGGGCACCTCCCACCTCCCAGAGGTTCGCCGTGACCACCGGGGAGCCCACG  298

Qy   1567 ProArgLeuGlnGluGlySerLeuSerLeuSerLysAlaSerGlnAsnArgLysLeuThr 1586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 CCCGCGCCTGCAGGAGGGCCAGCCTTTCGTCCAGCAAGCATCCCAGGAACCGAAAGCTGACG  238

Qy   1587 SerThrProArgGluIleAlaAlaLysSerProThrValProGluHisHisProHis 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 TCGACGCCCTCGTGAGATCGCCAAGTCCCCGACCGTGCCCGAGCACCACCCCACAC  178

Qy   1607 ProIleSerProTyrGluHisGluLeuArgGlyValSerGlyValAspLeuTyrArgSer 1626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 CCCATCTCGCCCTATGAGCACGAGCTGCGCGGGGTGAGTGGCGTGGACCTGATCGCAGC  118

Qy   1627 HisIleProLeuAlaPheAspProThrSerIleProArgGlyIleProLeuAspAlaAla 1646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    117 CACATCCCCCTGGCCTTCGACCCCACCTCCATACCCCGGGCATCCCTCGACGCAGCC  58

Qy   1647 AlaAlaTyrTyrLeuProArgHisLeuAlaPro 1657
          ||||||||||||||||||||||||||||||
Db     57 GCTGCCTACTACCTGCCCCGACACCTGGCCCC  25

RESULT 71
BG485444/c
LOCUS       BG485444                902 bp    mRNA          linear
DEFINITION  602505328F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618783 5',
            mRNA sequence.
ACCESSION   BG485444
```

```
Db      2 AGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGGGTGGTCACCCTGGCC  61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   2137 GlnHisIleSerGluValIleThrGlnAspTyrThrArgHisProGlnGlnLeuSer 2156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 CAGCACATCAGTGAGGTCATCACACAGGACTACACCCGCCACCCACAGCAGCTCAGC  121

Qy   2157 AlaProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeu 2176
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GCACCCCTGCCCGCCCCCCTTACTCCTTCCCTGGGGCCAGCTGCCCCGTCCTGGACCTC  181

Qy   2177 ArgArgProSerArgTyrLeuProTyrProProAspHisGlyAlaProAlaArgGly 2196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CGCCGCCCACCCAGTGACCTCTACCTCCGGCCCCGACCATGGTGCCCCGGCCCGTGGC  241

Qy   2197 SerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGly 2216
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 TCCCCCACAGCGAAGGGGCAAGAGTCTCCAGAGCCAAACAAGACTTCGGTCTTGGGT  301

Qy   2217 GlyGlyGluAspGlyIleGlGluProValSerProProGluGlyMetThrGluProGlyHis 2236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 GGTGGTGAGGACGTATTGAACCTGTGTCCCACCGGAGGGCATGACGGAGCCAGGGCAC  361

Qy   2237 SerArgSerAlaAlaValTyrProLeuLeuTyrArgAspGlyGluGlnThrGlnProSerArg 2256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TCCCGGAGTGCTGTGTACCCGCTGCTATACCGGGATGGGAACAGACAGCCAGCAGG  421

Qy   2257 MetGlySerLysSerProGlyAsnThrSerGlnInProProAlaPhePheSerLysLeuThr 2276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTTCAGCAAGCTGACC  481

Qy   2277 GluSerAsnSerAlaMetValLysSerLysLysSerGlnGluIleLeuAsnLysLeuAsnThr 2296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GAGAGCAACTCCCGCCATGGTCAAGTCCAAGAAGAGCAAGAGATCAACAAGAAGCTGAACACC  541

Qy   2297 HisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyIleThrGluIlePheAsnMet 2316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CACAACCGGAATGAGCCTGAATACAATATCAGCCAGCCTGGGACGGAGATCTTCAATATG  601

Qy   2317 ProAlaIleThrGlyThrGly 2323
          |||||||||||||||||||
Db    602 CCCGCCCATCACCGGGACAGGT  622

RESULT 70
BF530324/c
LOCUS       BF530324                657 bp    mRNA          linear     EST 11-DEC-2000
DEFINITION  602071630F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4214551
            5', mRNA sequence.
ACCESSION   BF530324
VERSION     BF530324.1  GI:11617698
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 657)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9788  row: 1  column: 08
            High quality sequence stop: 657.
FEATURES            Location/Qualifiers
     source         1..657
                    /organism="Homo sapiens"
```

```
VERSION      BG485444.1  GI:13417723
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 902)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1377 row: g   column: 08
             High quality sequence stop: 629.
FEATURES             Location/Qualifiers
     source          1..902
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:461878"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
                     SfiI (ggccgctcggcc); Site 2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3', and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb).  12/15 colonies contained inserts.
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
  Pred. No.:                  1.75e-36
  Score:                      1104.50
  Percent Similarity:         79.94%         Length:        902
  Best Local Similarity:      77.78%         Matches:       252
  Query Match:                8.36%          Conservative:  7
  DB:                         12             Mismatches:    31
                                             Indels:        36
                                             Gaps:          5

US-09-522-753-5 (1-2517) x BG485444 (1-902)

Qy  2103 LeuProHis--------LeuArgProLeuProGluSerGlnProSerSerProLeu 2119
             ||||||::::        |||||||||||     ||||
Db   890 CTACCAAACTGCGCCCGCCTGAGACCAGCCCT-------------TGTCCACCCGGG 846

Qy  2120 LeuGlnThrAlaProLeuTyrSerPheProGlyAlaSerCysProValLeu--AspLeuA 2177
             ||  |||    |||  ||| |||||||||||||      ||||||
Db   845 CTCCAGGCGCCGCCCCCAGGTCAAAGGTTCACCACCGGGTTCCACCTGGCCCAGCACAAT 786

Qy  2140 SerGluValIle-ThrGlnAspTyrThrArgHis-HisProGlnLeuSerAla---P 2158
             |||||||||    ::|||||||||||||  :::|||||||||||||||
Db   785 CATGAGGTCATCAACACAGGACTACAGCCGGCACCAACCCACAGCAGTCAGCGCAACCC 726

Qy  2158 roLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeu--AspLeuA 2177
             ||  |||        ||| |||||||||||
Db   725 CTGCCCCGCCCCCTCCGACGCGGCCGTGCCCCAGGGCCAGCTGCCCGCCGGGACCTGCCG 666

Qy  2177 rgArgProSerAsp-LeuTyr-LeuProProAsp-HisGlyAlaPro-AlaArg 2195
             ||:||||||       ||  :::||||||||||
Db   665 CCGCACAACCCAGTGACCGGCGTACGGTCCCGCCGGCGACCAATGGTGCCCCCGGGCCGT 606

Qy  2196 GlySerProHisSerGluGlyGlyGlyLysArgSerGluProGluLeuAsnLysThrSerValLeu 2215
             ||||
```

Right column:

```
Db   605 GGCTCCCCACAGCGCAAGG-GGCAAGAGGTCTCCAGAGCCAAACAAGAGACGTCGT--CTT 549
Qy  2216 GlyGlyGlyGluAspGlyIleGluLeuProValSerProProGluGlyMetThrGluProGly 2235
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   548 GGTGCTGTTGAGGACGTATTGAACCTGTGTCCCCACCGGAGGGCATGACGGAGCCAGGG 489
Qy  2236 HisSerArgSerAlaValTyrProLeuLeuTyrArgArgAspGlyGluGlnThrGluProSer 2255
             ||||||||||||||||||||||||||||||||||||||||||
Db   488 CACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCCGGATGGGGAACAGACGGAGCCCAGC 429
Qy  2256 ArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeu 2275
             |||||||||||||||||||||||||||||||||||||||||
Db   428 AGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCGCCAGCCTTCTTCAGCAAGCTG 369
Qy  2276 ThrGluSerAsnSerAlaMetValLysSerGlyLysLysGlnIleLeuAsnLysLeuAsn 2295
             ||||||||||||||||||||||||||||||||||||||||||
Db   368 ACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAGAGCAAGAGATCAACAAGAAGCTGAAC 309
Qy  2296 ThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlyThrGluIlePheAsn 2315
             |||||||||||
Db   308 ACCCACAACCGGAATGAGCCTGAATACAGC- 279
Qy  2316 MetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGlnAlaValGlnHisGluHis 2335
                        ------------------CTTATGACCTATAGAAGCCAGGCGGTGCAGGAACAT 243
Db   278 ---------------
Qy  2336 AlaSerThrAsnMetGlyLeuGluAlaIleAlaIleLeuArgLysAlaLeuMetGlyLysTyrTyrAsp 2355
             |||||||||||||||||||||||||||||||||||||||||
Db   242 GCCAGCACCAACATGGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGGGTAAATATGAC 183
Qy  2356 GlnTrpGluGluProProLeuSerArgAlaAsnAlaPheAsnProLeuAsnAlaSerAla 2375
             |||||||||||||||||||||||||||||||||||||||
Db   182 CAGTGGGAAGAGTCCCCGCGCTCAGCGCCAATGCTTTTAACCCCTGAATGCCAGTGCC 123
Qy  2376 SerLeuProAlaAlaMetProIleThrArgAlaAlaAspGlyArgSerAspHisThrLeuThr 2395
             |||||||||||||||||||||||
Db   122 AGCCTGCCCGCTGCTATGCCCATAACCGCTGCTGCGACGGAGTGACCACACTCACC 63
Qy  2396 SerProGlyGlyGlyProGlyGlyGlyGluGlyGlyGlyLysArgSerGluProGluLeuAsnLysThrSerValAlaLys 2415
             ||||||||||||||||||||||||||||||||||||||||||||
Db   62 TCGCCAGGTGGCGCGGGGAAGGCCAAGGTCTCTGCAGACCCAGCAGCGCCGAAAAGCCAAG 3

RESULT 72
CB216442
LOCUS       CB216442                 625 bp    mRNA    linear    EST 06-FEB-2003
DEFINITION  NISC_nq05b08.y1 NICHD_HS_Ut2 Homo sapiens cDNA clone IMAGE:5938818
            5', mRNA sequence.
ACCESSION   CB216442
VERSION     CB216442.1  GI:28264634
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM13166  row: C  column: 15
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1..625
                     /organism="Homo sapiens"
```

```
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="IMAGE:5938118"
          /sex="female"
          /tissue_type="normal endometrium, mid-secretory phase,
          cycle day 23"
          /lab_host="DH10B (T1-resistant)"
          /clone_lib="NICHD_HS_Ut?"
          /note="Organ: uterus; Vector: pCMV-SPORT6.1.ccdb (ResGen,
          Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV;
          Cloned unidirectionally from microquantity amounts of mRNA
          from normal endometrial tissue (mid-secretory phase, cycle
          day 23). Average insert size 1.6 kb. Library constructed
          by ResGen (Invitrogen Corporation)."

ORIGIN

Alignment Scores:
Pred. No.:             1.88e-36        Length:           625
Score:                 1099.50         Matches:          208
Percent Similarity:    99.52%          Conservative:     0
Best Local Similarity: 99.52%          Mismatches:       0
Query Match:           8.32%           Indels:           1
DB:                                    Gaps:             1

US-09-522-753-5 (1-2517) x CB216442 (1-625)

Qy   2163 LeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProProSerAsp 2182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCTACTCCTTCCCTGGGGCCAGCTGCGCCCGTCCTGGACCTTCCGCGGCCACCCAGTGAC 60

Qy   2183 LeuTyrLeuProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGly 2202
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTCTACCTCCCCGACCATGGTGCCCCCGGCTCGCCCGTGGCTCCCCCACAGCGAAGGG 120

Qy   2203 GlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIle 2222
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCAAGAGGTCTCCAGAGCCAAACAAGACGTCGGTCTTGGGTGGTGAGGACGGTATT 180

Qy   2223 GluProValSerProProGlyGluGlyMetThrGlyProGlyHisSerArgAlaValTyr 2242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GAACCTGTGTCCCCACCGGAGGGCATGACGGGACCAGGGCACTCCCGAGTGCTGTGTAC 240

Qy   2243 ProLeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerPro 2262
          |||||||||||||||||||||||||||||||||          |||||||||||||||||
Db    241 CCCCTGCTGTACCGGGATGGGGAACAGACAGAGCCC---AGGATGGGCTCCAAGTCTCCA 297

Qy   2263 GlyAsnThrSerGlnProAlaPheSerLysLeuAsnThrGluSerAsnSerAlaMet 2282
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 GGCAACACACCAGCCAGCCCGCCTTCTTCAGCAAGCTGAACCGAGAGCAACTCCGCCATG 357

Qy   2283 ValLysSerLysLysGlnIleLysAsnLysLysLeuAsnThrHisAsnArgAsnGluPro 2302
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 GTCAAGTCCAAGAAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGGAATGAGCCT 417

Qy   2303 GluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThr 2322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 GAATACAATATCAGCCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACCGGAACA 477

Qy   2323 GlyLeuMetThrTyrArgSerGlnAlaValGlnHisAlaSerThrAsnMetGlyLeu 2342
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    478 GGCCTTATGACCTATAGAAGCCAGGCGGTGCAGCACGCCAGCACCAACATGGGGCTG 537

Qy   2343 GluAlaIleIleArgLysAlaLeuMetGlyTyrLysTyrAspGlnTrpGluSerProPro 2362
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    538 GAGGCCATAATTAGAAAGGCACTCATGGGTAAATATGACCAGTGGGAAGAGTCCCCGCCG 597

Qy   2363 LeuSerAlaAsnAlaPheAsnProLeu 2371
          |||||||||||||||||||||||||||
Db    598 CTCAGCGCCAATGCTTTTAACCCTCTG 624

RESULT 73
BX732681
LOCUS       BX732681        825 bp     mRNA      linear      EST 18-NOV-2003
```

```
DEFINITION  BX732681 XGC-tadpole Silurana tropicalis cDNA clone TTpA071k07 5',
            mRNA sequence.
ACCESSION   BX732681
VERSION     BX732681.1  GI:38405422
KEYWORDS    EST.
SOURCE      Silurana tropicalis (western clawed frog)
  ORGANISM  Silurana tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
  TITLE     Sanger Xenopus tropicalis EST project 2001 (11_2003)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Croning MDR
            Sanger Institute
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: trop@sanger.ac.uk
            Sanger Xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: TTpA071k07.p1kSP6
            Sequencing_primer: SP6
            This sequence is from a Xenopus Gene Collection (XGC) library
            constructed by Nigel Garrett.
            cDNA was oligo dT primed from 5ug of poly A+ RNA from tadpole
            embryos. EcoRI-NotI cut cDNA was then ligated into pCS107 with
            EcoRI at the 5' end and NotI at the 3' end.
            Vector: pCS107; Site_1: EcoRI; Site_2: NotI
            Host: Escherichia coli DH10B.
FEATURES             Location/Qualifiers
     source          1..825
                     /organism="Silurana tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="TTpA071k07"
                     /dev_stage="tadpole (stage 35-40)"
                     /lab_host="E. coli DH10B"
                     /clone_lib="XGC-tadpole"
                     /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
                     was oligo dT primed from 5ug of poly A+ RNA from tadpole
                     embryos. EcoRI-NotI cut cDNA was then ligated into pCS107
                     with EcoRI at the 5' end and NotI at the 3' end"
ORIGIN

Alignment Scores:
Pred. No.:             2.81e-36        Length:           825
Score:                 1098.50         Matches:          216
Percent Similarity:    85.82%          Conservative:     20
Best Local Similarity: 78.55%          Mismatches:       38
Query Match:           8.31%           Indels:           1
DB:                                    Gaps:             1

US-09-522-753-5 (1-2517) x BX732681 (1-825)

Qy    138 AspArgSerLeuThrGlyLysLeuGluProValSerProProSerProProHisThrAsp 157
          ||||||       |||||  |  |||    |   |||||   ||  |    ||
Db      3 GACAGGGGTTGCCCAGCAAACATGAGACTACTTCGTCACCTTCTTCTGGGCAACCTGGA 62

Qy    158 ProGluLeuGluLeuValProProArgLeuSerLysGluLeuIleGlnAsnMetAsnAsp 177
          ||   ::|::|||    | |  ::: ||  ||||  | ||  ||| |::  |
Db     63 GAGGAACAGGAAGAGCCTCCATCAAACTTCTAAAGAGAGAACTATACAAAGTATGGAC 122

Qy    178 ArgValAspArgGluIleThrMetValGluGlnGlnIleSerLysLeuSerLeuLysGln 197
          ||  |  :|:|:   ||      |||||  ||  | |    ||    |  |::|
Db    123 CGCGTTGACCGCGAGATTGCTAAGTTGAACAGCAAATCTTAAAGTTGAAGAAAAAACAG 182

Qy    198 GlnGlnLeuGluGluGluProProGluAlaLysProProGlnProGluGluValSerProPro 217
          |  |||| |||    | |  ||    |  |:  ||   ||  |: ||
Db    183 CAACAGCTTGAAGAAGAAGCAGCAAAACCTCCAGAGCCAGAGAGCCCGTCTCCCCTCCA 242

Qy    218 ProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGluLysArgLysLys 237
          |  ::|||       |||  :| |  |  |||      |::|    | |::|||||
Db    243 CCTGTGGAACAGACAGCATCGCAGTAGTCCAAATTATTTATGATGAAAACCGGAAAAA 302
```

```
                              Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                              Average insert size 1.5 kb. Library prepared by Life
                              Technologies."

ORIGIN

Alignment Scores:
Pred. No.:          7.14e-36          Length:            922
Score:              1090.00           Matches:           236
Percent Similarity: 70.72%            Conservative:      8
Best Local Similarity: 68.41%         Mismatches:        25
Query Match:        8.25%             Indels:            77
DB:                 10                Gaps:              3

US-09-522-753-5 (1-2517) x BE542336 (1-922)

Qy    80 GlnGluLeuHisLeuArgProGluSerHisSerTyrLeuProGluGlyGlyLysSerGlu   99
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAGGAGCTCCACCTGCGGGCAGAGTCCCACTCATACCTGCCGGAGCTGGGGAAGTCAGAG   60

Qy   100 MetGluPheIleGluSerLysArgProArgLeuLeuLeuProAspProArgLeuLeuArg  119
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ATGGAGTTCATTGAAAGCAAGCGCCCTCGGCTGCTGCTCCTGCCCTGACCCCTGCCGCA  120

Qy   120 ProSerProLeuAlaAlaThrGlyGlnHisProAlaGlyAspLeuThrLysLysAspArg  139
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCGTCACCCCTGCTGGCCGCGGGCCAGCATCCGGGGATCTGAAGACCTCACCAAGGACCGT  180

Qy   140 SerLeuThrGlyLysGluProValSerProSerProProHisThrAspProGlu      159
         ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   181 AGCCTGACGGGCAAGCTGGAACCGGTGTCTCCCCAGGCCCCAGACACTGACCCTGAG  240

Qy   160 LeuGluLeuValProProArgLeuSerGlyGluLeuLeuGluLeuSerLysLysGlnGlnLeuGlnGln  179
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTGGAGCTGGTGCCGCCCCACGCGTGTCCAAGGAGGAGCTGATCCGAAGCATGGACCGCGTG  300

Qy   180 AspArgGluIleThrMetValGluGlnGlnIleLeuSerLysLysGlnGlnLeuGlnGln  199
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GACCGAGAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGCAGCAGAACAG  360

Qy   200 LeuGluGluGlnAlaAlaAlaLysProProGluProGluLysProValSerProProIle  219
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CTGGAGGAGGAGGCTGCCAAGCCCGGAGCCTGAGAAGCCCGTGTCACCCGCCGCCATC  420

Qy   220 GluSerLysHisArgHisGlnProGlnProSerLeuValGlnIleLeuTyrAspGluAsnArgLysLysAlaGlu  239
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAGTCGAAGCACCCCAGCCTGGTGCAGATCATCTACGACGAGAACCGG-------------  468

Qy   240 AlaAlaHisArgIleLeuGluGlyLeuGlyLeuProGlnValGluLeuProLeuuProLeuTyrAsnGln  259
         ||||||||||||||||||||||||||||||||||||||||||||::::||
Db   469 ------------------------------------------CCGGTGTACAACACCAG  483

Qy   260 ProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLys  279
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 GCCCTCCGACACCCGGCAGTATCATGAGAACATCAAAATAAAGCAGGCGGATGCGGAAGAAG  543

Qy   280 LeuIleLeuTyrPheLysLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys  299
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 CTAATCTTGTACTTCAAGAAGAGGAGAAATCACGCTCCGAAACAATGGAACAGAAGTTCTGC  603

Qy   300 GlnArgTyrAspGlnLeuMetGluMetGluLeuGluLeuLysLysLysValGluArgIleGluAsnAsn  319
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604 CAGGCGCTATGACCAGCTCATGGAGCGCTGGGAGAAGACCGTGGAGCGCATCGAGAACAAC  663

Qy   320 ProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluGlnPheProGlu      339
         ||| ||||||||||||||||||||||||
Db   664 CCCCG-CGGCGGCGGGCCAAGGAGAGC---------------------------------  686

Qy   340 IleArgLysGlnLysArgGlnLeuGlnArgMetGlnSerArgValGlyArgGlySer     359
         |||||||||||||||||||||||||||||||||
Db   686 ------------------------------------------------------------  686

Qy   360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu  379
```

```
Qy   238 AlaGluGluAlaAlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyr  257
         |||||||||||||||||||||||||::::|||||||||||||||||||||
Db   303 GCAGAGGAGGCACACATAAGATCTTAGAAGGTCTAGGTCCAAAAGTAGAGCTGCCTCTTTAC  362

Qy   258 AsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArg  277
         |||||||| |||||||||:::||
Db   363 AACCAGCCCTCAGACACAAAAGTTTATCATGAAAACATCAAGAACGAATCAGGTGATGAGG  422

Qy   278 LysLysLeuIleLeuTyrPheLysLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLys  297
         |||||||||||||:::|||||||||
Db   423 AAAAAACTAATACTGTCTTCTTAAGACAGAGAAAATCATGCTAGAAAAACTAAGGACACAGAAC  482

Qy   298 PheCysGlnArgTyrAspGlnLeuMetGluMetGluLeuGluLeuLysLysLysValGluArgIleGluGlu  317
         ||||||
Db   483 ATTTGTCAGCGCTATGATCAGCTGATGGAGCGCTTGGGAGGAAGAAGTTGACACGATAGAA  542

Qy   318 AsnAsnProArgArgArgAlaLysGluSerLysValArgGluTyrTyrGluLysGlnPhe  337
         ||||||||||||||
Db   543 AATAATCCTCGGAGAAAAGCAAAGGAGCAAAACCCGAGAATATTATGAGAAGCAATTT  602

Qy   338 ProGlnIleArgLysGlnLysArgGlnLeuGlnArgMetGlnSerArgArgValGlyGlnArg  357
         |||||||||||||
Db   603 CCTGAAATAAGGAAGCAGAGAGAGCAGCAGGAGCGTTTCCAG---AGGGTTGGGCAGAGA  659

Qy   358 GlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAsp  377
         ||||||||||||:::
Db   660 GGAGCTGGGCTGTGCAACCATAGCCAGGAGTGAACATGAAATATCTGAAATTATTGAT  719

Qy   378 GlyLeuSerGluGluGlnIleGlnAsnLeuGlyLysGlnMetArgGlnLeuAlaValIleProPro  397
         |||||||||||||||||
Db   720 GGTCTTTTCTGAACAGAGAATAATGAAAAGCAGATGCGCCAACTCTCAGTCATTCCACCC  779

Qy   398 MetLeuTyrAspAlaAspGlnArgIleLeuLysPheAsnMet  412
         ||:::||||||||::::||||
Db   780 ATGATGTTTGATGCCGGACGCANAGGAGGGTAAAATTCATTAATATG  824
```

```
RESULT 74
BE542336
LOCUS       BE542336                   922 bp    mRNA    linear    EST 09-AUG-2000
DEFINITION  601067079F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3453499 5',
            mRNA sequence.
ACCESSION   BE542336
VERSION     BE542336.1  GI:9770981
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 922)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8436  row: m  column: 20
            High quality sequence stop: 701.
FEATURES             Location/Qualifiers
     source          1..922
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3453499"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_10"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
```

```
Db   686 ------------------------------------------------------------ 686

Qy   380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnMetArgGlnLeuAlaValIleProProMetLeu 399
              ||||||||            ::   ::
Db   687 ------CCGGGGAACCTGGAGAACCTGATTTCGCCAGTGGGCCGGGGATCCGCCCCATGCTG 740

Qy   400 Tyr---AspAlaAspGlnArgIle-LysPheIleAsnMet-AsnGlyLeuMetAlaA 418
              :::||||||||          ||   :::|||||       ||||||
Db   741 GTACGCAACGCTGACCAGCACGCGCCTACAAGTGACTCCACATGTAACGGGGCAATGGCCA 800

Qy   418 8pProMetLys 421
              ||| ||
Db   801 AGCCCCGAAAG 811

RESULT 75
BX869832/c
LOCUS        BX869832                682 bp    mRNA        linear   EST 17-DEC-2003
DEFINITION   AGENAE Rainbow trout multi-tissues-normalized (tcbk)
             Oncorhynchus mykiss cDNA clone tcbk0022c.i.19 5prim, mRNA sequence.
ACCESSION    BX869832
VERSION      BX869832.1  GI:39993419
KEYWORDS     EST.
SOURCE       Oncorhynchus mykiss (rainbow trout)
ORGANISM     Oncorhynchus mykiss
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Euteleostei;
             Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE    1  (bases 1 to 682)
AUTHORS      Govoroun,M., Guiguen,Y. and Le Gac,F.
TITLE        Construction and primary characterization of normalized cDNA
             libraries in rainbow trout, Oncorhynchus mykiss
JOURNAL      Unpublished (2003)
COMMENT      Contact: Guiguen Y
             INRA - SCRIBE
             Campus de beaulieu, RENNES cedex, 35042, France
             Tel: 02.23.48.50.09
             Fax: 02.23.48.50.20
             Email: Yann.Guiguen@beaulieu.rennes.inra.fr
             Sequence cleaned of vector, adaptor and repetitions. Contact us
             at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
             sequence.
             Plate: 0022  row: i  column: 19
             Seq primer: M13R.
FEATURES             Location/Qualifiers
     source          1..682
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="tcbk0022c.i.19"
                     /tissue_type="multi-tissues"
                     /dev_stage="from embryos to adults"
                     /lab_host="DH10B"
                     /clone_lib="AGENAE Rainbow trout multi-tissues-normalized
                     (tcbk)"
                     /note="vector: pT7T3D-pac; Rainbow trout
                     multi-tissues-normalized + 2 subtractions; Clone
                     distribution : AGENAE Resource centre, Francois PIUMI,
                     Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
                     Etude du genome (LREG), Domaine de Vilvert, 78352,
                     Jouy-en-Josas cedex, FRANCE, +33 (0) 1.34.65.28.02, +33
                     (0) 1.34.65.22.73"
ORIGIN
Alignment Scores:
Pred. No.:              6.82e-36
Score:                  1087.00
Percent Similarity:     96.88%
Best Local Similarity:  91.96%
Query Match:            8.23%
DB:                     13

US-09-522-753-5 (1-2517) x BX869832 (1-682)
```

```
Qy   241 AlaHisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnPro 260
              ||||||||||||||||||||||||||||||||||||||::::||||||||||||||||||
Db   679 GCGCACAGGATTCTGGAAGGACTGGACCCCAGTAGAACTGCCTCTGTACACACCGCCA 620

Qy   261 SerAspThrArgGlnTyrHisGluAsnIleLysIleAsnGlnAlaMetArgLysLysLeu 280
              |||||||||||||||||::::||||||||||||||||||||||||||||||||||||||
Db   619 TCTGACACCAAGCAGTACCATGACAACATCAAAATAAACCAGCGATGGAGAAACTC 560

Qy   281 IleLeuTyrPheLysArgArgAsnHisAlaArgLysArgLysLysGlnTrpLysGlnPheCysGln 300
              ||||||||||||||::||:|||||||||||||||||||||||||||||||::||||||||||
Db   559 ATTCTTTACTTCAAGAGGAGAAACCATGTCTCCTAAGCAGTGGAACAGAAGTTCTGCCAG 500

Qy   301 ArgTyrAspGlnLeuMetGluAlaLeuGluLysLysLysValGluArgIleGluAsnAsnPro 320
              ||||||||||    |||||||::::||||||||||||||||||||||||||||||
Db   499 CGCTATGACCATCTGATGGAGGCCTGGAGGAGAAGAAGGTGGAGCGCATTGAGAACAACCCC 440

Qy   321 ArgArgArgAlaLysSerLysSerValArgGluTyrTyrGluValArgGluProGluIle 340
              |||||||||||||||||||||||||||||::||||||||||||||||||::|||
Db   439 CGACGCCGGGCTAAGGAGAGCAAGGTGCGAGAGTACTACGAGAAACAGTTCCCTGAGATC 380

Qy   341 ArgLysGlnArgGluGlnLeuGlnLeuArgMetGlnSerArgValGlyValArgGlySerGly 360
              |||||||||||||||||::||||||||||||||||||||||||||||||||::|||
Db   379 CGCAAGCAGAGAGAGCTGCAGGAGCAGCTGCAGAGCCGGGCCAGCCGTGAGGAGGA 320

Qy   361 LeuSerMetSerAlaAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSer 380
              |||::::   ||||||||||||||||::::||||||||||||||||:::||||
Db   319 CTGGCCTCATCTGCAGCCCGCCAGTGAACACGAGGTCTCTGAGATCGTTGACGGCATATCA 260

Qy   381 GluGlnAsnLeuGluLysGlnMetArgGlnLeuLeuAlaValIleProProMetLeuTyr 400
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259 GAGCAAGAGAACACCGAGAAGCAAATGCCACTGGCGGTCATCCCTCCCATGTTGTTT 200

Qy   401 AspAlaAspGlnArgIleLysPheIleLysAsnMetAsnGlyLeuMetAlaAspProMet 420
              ||||||||:::||||||||||||||||||||||||||||||||||||
Db   199 GACGCCGAGCAGCAGCCGCATCAAGTTCATCAACATGACGGCCTGATGACGACCCCATG 140

Qy   421 LysValTyrLysAspArgGlnValMetAsnMetTrpSerGluGlnGlnLysGluThrPhe 440
              |||||||||||||||||||||||||||||||||||||:::|||
Db   139 AAGGTGTACAAGGACCGGCAGGTCATGAACATGTGGAGCAGAGGAGAAGGAGACACCTTC 80

Qy   441 ArgGluLysPheMetGlnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArg 460
              ||||||||:::||||||||||||||||||||||||||||||||||||||
Db   79  CGCCAGAAGTTCATCCAGCACCCAAAAAACTTTGGCCTGATTGCTTCTTTCCTGGAGAGA 20

Qy   461 LysThrValAla 464
              |||||||||||
Db   19  AAGACGGTGGCC 8

RESULT 76
BI081252
LOCUS        BI081252                790 bp    mRNA        linear   EST 20-JUN-2001
DEFINITION   602879279F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5011003 5',
             mRNA sequence.
ACCESSION    BI081252
VERSION      BI081252.1  GI:14499582
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 790)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: The I.M.A.G.E. Consortium, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
```

```
              Plate: L1AM11059 row: m column: 20
              High quality sequence stop: 624.
FEATURES
     source       Location/Qualifiers
              1..790
                   /organism="Mus musculus"
                   /mol_type="mRNA"
                   /strain="FVB/N-3"
                   /db_xref="taxon:10090"
                   /clone="IMAGE:5011003"
                   /tissue_type="tumor, biopsy sample"
                   /dev_stage="5 months"
                   /lab_host="DH10B"
                   /clone_lib="NCI_CGAP_Mam2"
                   /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                   Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                   Library constructed by Life Technologies. Investigator
                   providing samples: Gilbert Smith, NIH"

ORIGIN

Alignment Scores:
Pred. No.:          9.7e-36        Length:         790
Score:             1085.00         Matches:        230
Percent Similarity: 77.48%         Conservative:   4
Best Local Similarity: 76.16%      Mismatches:     15
Query Match:        8.21%          Indels:         54
DB:                 12             Gaps:           3

US-09-522-753-5 (1-2517) x BI081252 (1-790)

Qy  2222 IleGluProValSerProProGluGlyMetThrGluProGlyHisSerArgSerAlaVal 2241
             ||||||||||||||||||||||||||||||||||||||||||||||:::||
Db     1 ATTGGAACCTGTGTC--CCACCAGAGGCATGACTGAGCCAGGACATGCTCGGAGCACTGCG  59

Qy  2242 TyrProLeuLeuTyrArgAspGlyGlnThrGlyProSerArgMetGlySerLysSer 2261
             |||||||||||||||||||||||   ||||||   |||||||||||
Db    60 TACCCACTGCTGTATCGAGACGGGAACAGGGCGGCCC---AGGATGGGCTCTAAGTCT 116

Qy  2262 ProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAla 2281
             |||||||||||||||||||||:||:||||||||||||||||||||||||||
Db   117 CCAGGCAACACCAGCCAGCCGCCAGCCTTCTCACTAAGTCACTGAGGAGCAACTCCGCC 176

Qy  2282 MetValLysSerLysSerGlnGluIleAsnLysLeuGluAsnThrHisAsnArgAspGlu 2301
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 ATGGTGAAGTCGAAGAAGCAGGAGATCAACAAGAAACTCAACACCCACAACCGGAACGAG 236

Qy  2302 ProGluTyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGly 2321
             ||||||||||||   |||||||||||||||||||||||||||||||||||| |||||||
Db   237 CCAGAATACAATATTGGCCAGCCTGGGACGGAAAATCTTCAACATGCCCGCCATCACTGGA 296

Qy  2322 ThrGlyLeuMetThrTyrArgSerGlnAlaValGlnHisGluHisAlaSerThrAsnMetGly 2341
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 GCAGGCCCTTATGCACCTGTAGAAGCAGGCGGTGCAAGAACACGCCAGCCACCAACATGGGG 356

Qy  2342 LeuGluAlaIleIleArgLysAlaLeuMetGlyLysMetGlyTyrAspTyrAspGlnTrpGluSerPro 2361
             ||||||||||||||||||||||||
Db   357 CTAGAGGCCCATTATTAGGAAAGGCACTCATG------------------------------ 386

Qy  2362 ProLeuSerAlaAsnAlaPheAsnProLeuAsnAlaSerLeuProAlaAlaMet 2381
             |
Db   386 ------------------------------------------------------ 386

Qy  2382 ProIleThrAlaAlaAspArgSerAspHisThrLeuThrSerProGlyGlyGly 2401
                                                         |||||||||
Db   387 ------------------------------------------GGTGAGGTGGG 398

Qy  2402 LysAlaLysValSerGlyArgProSerSerArgAlaLysSerProAlaProGlyLeu 2421
             ||||||||||||||||||||||||||||||||||||||||
Db   399 AAAGCCAAGGTCTTCTGCAGACCTAGCAGCCGAAAAGCCAAGTCGCCACCAGGCCTA 458

Qy  2422 AlaSerArgGlyAspArgProProSerValSerSerValHisSerGluGluValGlyAspCysAsnArg 2441
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   459 GGCGTCCGGAGACCGACCCCCTTCTGTCTCCTCAGTACACTCAGGAGGGGACTGCAATCGC 518
```

```
Qy  2442 ArgThrProLeuThrAsnArgValTrpGluAspArgProSerSerAlaGlySerThrPro 2461
             ||||||||||||::||||:||||||||||||||||||||||||||||||
Db   519 CGAACACCAATCACCAACCGTGTGGGAGGACCGCCCTCATCTGCAGGGGTCCACGCCA 578

Qy  2462 PheProTyrAsnProLeuIleMetCArgLeuGlnLeuAlaGlyValMetAlaSerPro---Pro 2480
             ||||||||||||||||||||||||||||||||||
Db   579 TTCCCTACAAACCTTGATTATGAGGCTACAGGCAGGTGTCATGGCCTCCACCAGAACA 638

Qy  2481 ProProGlyLeuProAlaGly-SerGlyProLeuAlaGlyProHis-HisAlaTrpAsp- 2499
             ||||||   ||||||||||||  |||||||||||||||||||||||
Db   639 CAAACTGGCCTTGCCGCAGCCAAGCGGGCCCTAGCTGGTCCCCACACACCCCTGGGACT 698

Qy  2500 GluGluProLysProLeu--LeuCysSerGlnTyrGluThrLeuSerAspSerGlu 2517
             ||||||||||||||||||||||||||||||||||   :::
Db   699 GAGGAGCCCAGCCCCTGGATGTTGTTCACGTATGAGACCTTCTGGACAAGCAG 754

RESULT 77
CA325207
LOCUS         CA325207                729 bp    mRNA    linear   EST 09-JUL-2003
DEFINITION    UI-M-FY0-ccn-b-19-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
              IMAGE:6821780 5', mRNA sequence.
ACCESSION     CA325207
VERSION       CA325207.1  GI:2454330S
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 729)
REFERENCE
  AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
  TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL     Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. Jim Lin, University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)
              Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..729
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6821780"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FY0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
                     Site 2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGCGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Alignment Scores:
```

Pred. No.:
Score: 1.3e-35
Percent Similarity: 1081.00
Best Local Similarity: 92.80%
Query Match: 88.14%
DB: 8.18%
14

Length: 729
Matches: 208
Conservative: 11
Mismatches: 16
Indels: 2
Gaps: 0

US-09-522-753-5 (1-2517) x CA325207 (1-729)

```
Qy  1402  LeuGlyProLeuLysLeuLysProAlaHisGluProAlaHisGluGlyLeuValAlaThrValLysGluAla  1421
          ||||||||||||||||||||||||||||||||||||:::||  ||||||||||||||
Db    23  CTGGGTCCCCTGAAGCTGAAGCCGACTCACGAGGGTCGTGGTAGC-ACTGTGAAGGAGGCG  81

Qy  1422  GlyArgSerIleHisGluIleProArgGluGluArgHisThrProGluLeuPro Leu  1441
          ||||||||||||||:::||||||||||||||||||||||
Db    82  GGCCGGCTCTATCATCATGAGATCCCGAGAGAGGAGCTGCGCCGCACACCTGAGCTACCCCTG  141

Qy  1442  AlaProArgProLeuLysGluGlyLySerIleThrGlnGlyThrProLeuGlyTyrAspThr  1461
          ||||||||||||||||||||||:::||||||||||||||:::
Db   142  GCACCACGGCCCTCGAAGAAGGAGGGTTCCATCACCCAGGGCACCCCACTCAAGTACGACTCT  201

Qy  1462  GlyAlaSerThrThrGlySerLysHisGlySerLeuIleGlySerProGly  1481
          ||||||   ||:::||||||||||||||||||||||||
Db   202  GGGGCAAGTACCACTGGCACCAAGAAACACGACGTGCGCTCCATCATCGGCAGCCCCGGC  261

Qy  1482  ArgThrPheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArg  1501
          :::||||||||||||||||||:::||||||||||||||
Db   262  CGGCCCTTCCCTGCCCTGCACCCGCTGGACATAATGGCTGACGCCCGGGCACTGGAGCGT  321

Qy  1502  AlaCysTyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerSerGlyGlySer  1521
          ||||||||||||||||||||  :::||||||||||||||
Db   322  GCCTGCTATGAAGAAGAGTCTGAAGAGCCGGTTCAGGAGACCAGCAGTGTGCAGGGGGCTCC  381

Qy  1522  IleAlaArgGlyAlaProValIleValProGlyLeuGlyLysProArgGlnSerProLeu  1541
          ||||||||||||:::|||||||||||||||||||||||
Db   382  ATCACACGTGGGGCTCCAGTCGTCGTGCCTGGACTGGGCAAGCCACGGCAAAGCCCACTG  441

Qy  1542  ThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThr  1561
          ||||||||||||||||||||||||||||||||||||
Db   442  ACTTACGGAGACCACGGGGCACCCTTCACCAGTGCCCAGTGGTCTCCCCTGTGACC  501

Qy  1562  MetArgGluProThrProArgLeuGlnGluGlySerLeuSerSerLysAlaSerGln  1581
          ||||||||||||||||||||||||||||||||||||
Db   502  ACGAGGGAGCCCACGCCACGCCTTCAGGAAGGCAGCCTCCATCCAGCAAGGCGTCCCAG  561

Qy  1582  AspArgLysLeuThrSerThrProArg-GluIleAlaLysSerProHisSerThrValPr  1601
          ||||||   :::|||||||||||||||||||||||
Db   562  GACCGGAAGCTGACATCTTACACCCGGNGAGATCGCCAAGTCCCCAAGTCCCCACACCAGCACTGTGCC  621

Qy  1601  oGluHisHisProHisProIleSerProTyrGluHisLeuLeuArgGlyValSerGlyVa  1621
          ||||||||||||||||:::|||||||||||||||||||
Db   622  CGAGCACCACCCTCACCCCATCTCCCCCTATGAGCACTTGCTCCGGGGCGTGACTGGTGT  681

Qy  1621  lAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSer  1636
          ||||||||   |||||||||
Db   682  GGACCTGTACCGTGGTCACATCCCCATTNGGCTTTGACCCCACCTCC  727
```

RESULT 78
BU475754
LOCUS   BU475754   809 bp   mRNA   linear   EST 30-NOV-2002
DEFINITION   603471578F1 CSEQRBN22 Gallus gallus cDNA clone ChEST348120 5`, mRNA
             sequence.
ACCESSION   BU475754
VERSION   BU475754.1   GI:25969331
KEYWORDS   EST.
SOURCE   Gallus gallus (chicken)
ORGANISM   Gallus gallus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
             Phasianinae; Gallus.
REFERENCE   1   (bases 1 to 809)
AUTHORS   Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
             Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.

FEATURES   Location/Qualifiers
     source   1..809
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Layer and broiler"
                     /db_xref="taxon:9031"
                     /clone="ChEST348120"
                     /sex="Male and female"
                     /tissue_type="Chondrocytes isolated from growth plate
                     cartilage"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="CSEQRBN22"
                     /note="Vector: pBluescript II KS(+); Site 1: EcoRI;
                     Site 2: NotI; This normalized library was constructed from
                     1 million independent clones. cDNA synthesis was initiated
                     using an oligo(dT) primer, using methylated C in the first
                     strand synthesis reaction. Following this first strand
                     reaction, double-stranded cDNA was blunted, ligated to
                     NotI adapters, digested with EcoRI, size-selected, and
                     cloned into the NotI and EcoRI compatible sites of a
                     custom modified MCS of the pBluescript (KS+) vector. The
                     library was normalized in 2 rounds using conditions
                     adapted from Soares et al. PNAS (1994) 91: 9228-9232 and
                     Bonaldo et al., Genome Research 6 (1996): 791, except that
                     a significantly longer reannealing hybridization was
                     used."

ORIGIN

Alignment Scores:
Pred. No.:
Score: 2.84e-35
Percent Similarity: 1074.00
Best Local Similarity: 85.87%
Query Match: 81.27%
DB: 8.13%
                            13

Length: 809
Matches: 230
Conservative: 13
Mismatches: 17
Indels: 25
Gaps: 4

US-09-522-753-5 (1-2517) x BU475754 (1-809)

```
Qy   334  GluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnArgMetGln------  351
          ||||||||||||||||||||||||||||||||||||
Db     4  GGAGAAGCAATTCCCTGAGATCCGAGACAGCAGCGGGAACTGCAAGAGCGCATGCAGAGGTTA  63

Qy   352  ------SerArgValGlyArgGlySer---GlyLeuSerMetSerAlaAlaArgSer  368
                |||||||||||||||||||||||
Db    64  GAGGAGAAGCCGGGTAGGACAGAGGGGCACGCGGAGGGCCTCCATGTCGGCTGCCGCGCAGT  123

Qy   369  GluHisGluValSerGluIleIleAspGlyLeuSerGluGlnLeuGlyLysGluGlyLysGln  388
          |||||||||||||||||||||||||||||||||||||||
Db   124  GAGCACGAGGTGTCGGAGATCATCGATGGCCTCAGAGAACAAGGAGAAGCAG  183

Qy   389  MetArgGlnLeuAlaValAlaValIleProProMetLeuTyrArgAspAlaAspProIleLys  408
          |||||||||||||||||||||||||||||||||||||
Db   184  ATGCGCCAGCTCGCCGTCATCCCACCCATGCTCTATAGATGATCGTGACCACCAGCAGCCATTAA  243

Qy   409  PheIleAsnMetAsnGlyLeuMetAlaAspProMetLysValTyrLysAspArgGlnVal  428
          ||||||||||||||||||||||||||||||||||||
Db   244  TTCATCAACATGAACGGCCTCATGGACGACCCCATGAAGGTCTCAAAGGACCGGCAGGTG  303

Qy   429  MetAsnMetTrpSerGluGluGluThrPheArgGluLysPheMetGlnHisPro  448
          ||||||||||||||||||||||||||||||||||||
```

```
Db   304 ATGAACATGTGGAGCGAGCGAGGAGAAGGAGACCTTCCGGGAAAAGTTCATGCGCACCCCA 363

Qy   449 LysAsnPheGlyLeuIleAlaSer-PheLeuGluArgLysThrValAlaGluCysValLe 468
             ||||||||||||||     |||||||:::||||||  ||||||||:::|||||
Db   364 AAGAACTTTGGCCTGATTGCCTCCTTTTTCTGGACAGGAAGACGGTGGCAGACTGCGTCCT 423

Qy   468 uTyrTyrTyrLeuThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrAr 488
             |||||||||||||||:::||||||||||:::|||||||
Db   424 GTACTACTACCTGACCAAGAAGAATGAGAACTACAAGAACCTCGTGAGGAGGAATTACCG 483

Qy   488 gArgArgGlyLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 508
             ||||||||||||||||||
Db   484 GCGCGGTGGGAAAAGCCAGCAGCAGCAA------------------------------ 514

Qy   508 nGlnGlnProMetProArgSerGlnGlnGluLysAsp---------GluLysGluLy 525
             |||||||||||||||||||       ||:::  :::
Db   515 ------------ATGCCCCGAGCAGCCAAGAAGAGAAAGATGAAACAAGAGAAAGAAACAG 564

Qy   525 sGluLysGluAlaLysLysGluGluLysLysProGluValGluAsnAspLysGluAspLe 545
             :::          :::||||||||||:::|
Db   565 GAGAAGAGGGGAGACAAGAAGAGAAGACAAGAAACAGAAAATGACAAGAAGAGAACT 624

Qy   545 uLeuLysGluLysGlyThrArgAspAspThrSer-GlyGluAspAsnAspGlyLysGluAlaValA 565
             |   ||:::|  |||:::::||||||||||||
Db   625 GACAAAGGACAAGAACGACCACGTCCGGGGAGGAACAACGATGAGGAAGAAACCGTCA 684

Qy   565 laSerLysGlyLysThrAlaAsnSerGlnGlyArgGlyArgArgIleThrArgS 585
             |||||||||||||||||||      :::
Db   685 CCTCCAAAGGCCGCAAAACCGC-AACAGCAGGGGCCGGCAAGGGGCATCACGGGCT 743

Qy   585 erMetAlaAsnGluAlaAsnSerGluGluAlaIleThrProGlnSerArgAlaGluLeuA 605
             |||||||||||||||:::||     ||||||||||||||
Db   744 CCATGGCCAACGAGGCCAACAGCGAGGAGGCCGGC-ACGCCGCAGCAGCGCCCGAGCTGG 802

Qy   605 laSer 606
             |||||
Db.  803 CCTCT 807
```

```
RESULT 79
BM011980
LOCUS       BM011980         788 bp   mRNA    linear   EST 30-OCT-2001
DEFINITION  603636389F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5458487 5',
            mRNA sequence.
ACCESSION   BM011980
VERSION     BM011980.1   GI:16526334
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1960 row: b  column: 24
            High quality sequence stop: 640.
FEATURES             Location/Qualifiers
     source          1..788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5458487"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
```

/clone_lib="NIH_MGC_47"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. DIrectionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:   2.9e-35   Length: 788
Score: 1073.50   Matches: 215
Percent Similarity: 95.20%   Conservative: 3
Best Local Similarity: 93.89%   Mismatches: 9
Query Match: 8.12%   Indels: 4
DB: 12   Gaps: 1

US-09-522-753-5 (1-2517) x BM011980 (1-788)

```
Qy  2291 AsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGly 2310
              |||||||||||||||||||||||||||||||||||||||||||
Db     3 AACAAGAAGCTGAACACCCACAACAGGAATGAGCCTGAATACAATATCAGCCAGCCTGGG 62

Qy  2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLysLysLysMetThrTyrArgSerGln 2330
              |||||||||||||||||||||||||||||||||||||||||||
Db    63 ACCGAGATCTTCAATATGCCCGCCATCACCGGACACAGGCCTTATGACCTATAGAAGCCAG 122

Qy  2331 AlaValGlnGluHisAlaSerThrAsnMetGlyValGluLeuGluAlaIleIleArgLysAlaLeu 2350
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 GCGGTGCAGGAACATGCCAGCACCAACATGGGCGTGGAGCTGGAGGCCATAATTAGAAAGCACTC 182

Qy  2351 MetGlyLysTyrAspGlnTrpGluSerProProLeuSerProAlaAsnAlaPheAsnProPro 2370
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 ATGGGTAAATATGACCAGTGGGAGAGTCCCCCGCTCAGCCCGGCCAATGCTTTTAACCCT 242

Qy  2371 LeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSer 2390
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CTGAATGCCAGTGCCAGCCTGCCCGCTGCTATGCCCATAACCGCTGCTGACGGACGGAGT 302

Qy  2391 AspHisThrLeuThrSerProGlyGlyLysLysAlaLysValSerGlyArgProSer 2410
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 GACCACACCACTCACCTCGCCAGGTGGCGGGAAGGCCAAGGTCTCTGGACACCCAGC 362

Qy  2411 SerArgLysAlaLysSerProAlaProGlyLeuAlaSerGlyArgAspArgProSerVal 2430
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 AGCCGAAAAGCCAAGTCCCCGGCCCCCGGGCCTGGCATCTGGGAGACCGGCCACCCCTCGTC 422

Qy  2431 SerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsnArgValTrp 2450
              ||||||||||||||||||||||||||||||||||||||||||||||:::|
Db   423 TCCTCAGTGCACTCGGAGGGAGACTGCAACCGCCGGACGCCGCTCACCAACCGCGTGTGG 482

Qy  2451 GluAspArgProSerSerAlaGlySerThrProPheProTyrAsnProLeuIleMetArg 2470
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483 GAGGACAGGCCCCTCGTCGGCAGTTCCACGCCATTCCCTACAAGCCCCTGATCATGCGG 542

Qy  2471 LeuGlnAlaGlyValMetAlaSerProProProGlyLeuProAlaGlySerGlyPro 2490
              ||||||||||||||||||||||||||||||||||||||||||:::|
Db   543 CTGCAGGCGGGGGTGTCATGGCTTCCCCATCCCAAC-GGGCTCCCGGCGGGCAACCGGCCC 601

Qy  2491 LeuAlaGlyLeuProHisHisAlaTrpAspGluProLysProLeu---LeuCysSerGln 2509
              ||||||||||||||||||||||||||||||:::|||::::|  ||:::
Db   602 CTGCCTGCCCCACA-CACGCCTGGACGGAGAGCCCAAGCCCCATGGCTCTGGCTCGCAAG 660

Qy  2510 TyrGlu-ThrLeuSerAspSerGlu 2517
              ||||| |||||:::  ||||||
Db   661 TACGAAGACATTCTCCGACAGCGAA 685
```

```
RESULT 80
BB391091
LOCUS       BB391091         688 bp   mRNA    linear   EST 21-JUL-2000
```

DEFINITION   601286040F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3607929 5',
              mRNA sequence.
ACCESSION   BE391091
VERSION   BE391091.1 GI:9336456
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 688)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM259 row: p column: 10
              High quality sequence start: 8
              High quality sequence stop: 684.
FEATURES   Location/Qualifiers
     source   1..688
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3607929"
                     /tissue_type="endometrium, adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_44"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.:   3.32e-35   Length: 688
Score: 1070.50   Matches: 222
Percent Similarity: 98.24%   Conservative: 1
Best Local Similarity: 97.80%   Mismatches: 4
Query Match: 8.10%   Indels: 4
DB: 10   Gaps: 0

US-09-522-753-5 (1-2517) x BE391091 (1-688)

Qy   82 LeuHisLeuArgProGluSerHisSerTyrLeuProGluLeuGlyLysSerGluMetGlu 101
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6 CTCCACCTGCGGCCAGAGTCCCACTCATACCTGCCCAGCTGGGAAGTCAGAGATGGAG 65

Qy 102 PheIleGluSerLysArgProArgLeuGluLeuLeuProAspProLeuLeuArgProSer 121
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66 TTCATTGAAAGCAAGCGCCCTCGGCTAGAGCTGCTGCTCCCCCCTGCTGCGACCGTCA 125

Qy 122 ProLeuLeuAlaThrGlyGlnProAlaGlySerGluAspThrLysAspArgSerLeu 141
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 126 CCCCTGCTGCCACGGCCAGCCTGCCGGGATCTGAAGACCTACCAAGGACCCTAGCCTG 185

Qy 142 ThrGlyLysLeuGluProValSerProProSerProProHisThrAspProGluLeuGlu 161
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 186 ACGGGCAAGCTGGAACCGGTGTCTCCCCCCAGCCCCCCGCACACTGACCCTGAGCTGGAG 245

Qy 162 LeuValProProArgLeuSerLysLysGluGluLeuGlnIleAsnMetAspArgValAspArg 181
        |||||||||||||||||||||||||||::|||||||||||||||||||:|||||||||
Db 246 CTGGTGCCGCCACGGCTGTCCAAGAAGGAGGAGCTGCAGATCAATATGGACCGA 305

Qy 182 GluIleThrMetValGluGlnGlnIleSerLeuLysLeuLysLysGlnGlnLeuGlu 201
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 306 GAGATCACCATGGTAGAGCAGCAGATCTCTAAGCTGAAGAAGCAGCAACAGCTGGAG 365

Qy 202 GluGluAlaLysProGluProGluLysProGluLysProValSerProValSerGluSer 221
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 366 GAGGAGGCTGCCAAGCCGCCCAGCCTGAGAAGCCCGTCACCCGCCATCGAGTCG 425

Qy 222 LysHisArgSerLeuValGlnIleIleTyrArgAspGluAsnArgLysLysAlaGluAlaAla 241
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 426 AAGCACCGCAGCCTGGTGCAGATCATCTACCGAGACCCGAAGAGGCCTGAAGCTGCA 485

Qy 242 HisArgIleLeuGluGlyLeuGlyProGlnValGluLeuProLeuTyrAsnGlnProSer 261
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 486 CATCGGATTCTGGAAGGCCTGGGCCCCCAGGTGGAGCTGCCGCTGTACAACCAGCCCTCC 545

Qy 262 AspThrArgGlnIleTyrHisGluAsnIleLysIleAsnGlnAlaMetArgGlyLysLeuIle 281
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 546 GACACCCCGCCAGTATCATGAGAACATCAAATA-AACCAGGCGATGCGGAAGAAGCTAATC 604

Qy 282 LeuTyrPheLysArgArgAsnHisAlaAlaArgLysGlnTrpLysPheCysGlnArg 301
        |||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db 605 TTGTACTTCAAGAGGAGGAT-CACGCTCGGAAACA-TGGGACCAGAGTC-TGCCAGCGC 661

Qy 302 TyrAspGlnLeuMetGluAla 308
        |||||||||||||||||||||
Db 662 TATGACCAGCTCATGGAGGCT 682

RESULT 81
BF058781/c
LOCUS   BF058781   664 bp   mRNA   linear   EST 16-OCT-2000
DEFINITION   7k34g07.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3477325 3'
              similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
              ; mRNA sequence.
ACCESSION   BF058781
VERSION   BF058781.1 GI:10812677
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 664)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
              R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
              Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
              I.M.A.G.E. Consortium DNA Sequencing by: Washington University
              Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL, send email to:
              info@image.llnl.gov
              Seq primer: -40UP from Gibco
              High quality sequence stop: 467.
FEATURES   Location/Qualifiers
     source   1..664
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3477325"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Ov18"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCGGCGGCACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not

AUTHORS   Boardman,P.B., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
           Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE   A Comprehensive Collection of Chicken cDNAs
JOURNAL   Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE   22335534
 PUBMED   12445392
COMMENT   Contact: Simon Hubbard
           Department of Biomolecular Sciences
           University of Manchester Institute of Science and Technology
           (UMIST)
           PO Box 88, Manchester, M60 1QD, UK
           Tel: 01612008930
           Fax: 01612360409
           Email: Simon.Hubbard@umist.ac.uk.
FEATURES   Location/Qualifiers
     source   1..1088
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="White Leghorn, Hisex"
                     /db_xref="taxon:9031"
                     /clone="ChEST61g22"
                     /dev_stage="36"
                     /lab_host="DH10B"
                     /clone_lib="CSEQCHL12"
                     /note="Organ: heart; Vector: pBluescript II KS(+); Site_1:
                     EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
                     [Stratagene] vector to accommodate cDNA produced with the
                     T-trimmed protocol (Construction of uni-directionally
                     cloned cDNA libraries from messenger RNA for improved 3'
                     end DNA sequencing by Glenn Fu, et al. U.S. Patent #
                     6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
                     Ligate in double stranded adaptor containing BsgI and
                     BamHI sites [5'gcccggtgcagacccgatccgaaaaaag]
                     [5'aatcctttttctggatccgggggctgcacgc]"
ORIGIN

Alignment Scores:
Pred. No.:   1.18e-34   Length: 1088
Score: 1062.50   Matches: 235
Percent Similarity: 74.79%   Conservative: 35
Best Local Similarity: 65.10%   Mismatches: 64
Query Match: 8.04%   Indels: 29
DB: 13   Gaps: 9

US-09-522-753-5 (1-2517) x BU108220 (1-1088)

Qy 1117 ValLeuGluArgGlnIleGlyAlaIleSerGlnGlyMetSerValGlnLeuHisValPro 1136
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1 TTGCTGGAGAGACCTGTTGGCTCCATCTCCCAGGAATGCCATCCAGCTCCACACACCC   60

Qy 1137 Tyr---SerGluHisAlaAlaLysAlaProValGlyProValThrMetGlyValLeuProLeuPro 1155
          ::: ||| ||||||||||||||||||||||||||||||||||||||||
Db   61 TACAGCTCTGAGATGCAAAAGTTCCTGTCGGCTCCATCACCATGGGCGTCCACTGACA   120

Qy 1156 MetAspProLysLeuAlaProPheSerGlyValLysGlnGluGlnLeuSerProArg 1175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATGGATCCCAAGAACTGTGTGCCCTTTTCCTGGAGTAAAGCAGGAACAACTCTCCCAGA   180

Qy 1176 GlyGlnAlaGlyProGluSerLeuGlyValProThrAlaGlnGluAlaSerValLeu 1195
               ::: |||||||||||||||||||||||||||||||||||||||||
Db   181 AGCCAGGCCAGCCAGCCCTGGAAAGCCCTGGTGATGCAGCCCGGCCCCCGGAGGGCTCCGTCCTG 240

Qy 1196 ArgGlyThrAlaLeuGlySerValProGlyGlySerIleThrLysGlyIleProSerThr 1215
          |||||||||| |||:::::::||||||||||||||||::::::||
Db   241 CGGGGTAGCTCCCTCAGCTCTGCCTCAGCAGGAGCATCACCAAGGGACACCCACATCC   300

Qy 1216 ArgValProSerAspSerProArgAlaIleThrTyrArgGlySerIleThrHisGlyThrProAla 1235
          ||| ||| :::||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGACCCCTCCAGATGCCCATCAGCTACCGACCGAGGCTCCATCACCACGGCACCCGGCA   360

Qy 1236 AspValLeuTyrLysGlyThrHisThrArgIleIleGluGlyGluAspSerProSerArgLeu 1255
          ::: ||||||| ||||||||||||||:::|||||||||||||||||||||||
Db   361 GAAGTGCTGTACAAAGGAACCATTACCAGGATAATCGGAGAAGAGCAGCCCCAGCAGAGCG 420

I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Alignment Scores:
Pred. No.:   3.35e-35   Length: 664
Score: 1070.00   Matches: 201
Percent Similarity: 98.07%   Conservative: 3
Best Local Similarity: 97.10%   Mismatches: 4
Query Match: 8.10%   Indels: 0
DB: 10   Gaps: 0

US-09-522-753-5 (1-2517) x BF058781 (1-664)

Qy 1902 SerThrSerSerProValArgProAlaAlaThrPheProProAlaThrHisCysProLeu 1921
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   622 TCACCTTCATCACCTGTTCGCCAGGCTGCCACATTCCCCCTTGCCACCCACTGTCCCACTG 563

Qy 1922 GlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuLeuProLysGlu 1941
          ||||||||::::|| ||||||||||||||||||||||||||||::::||
Db   562 ACAGGCACCATGCGATGGGGTCTCACCCTCATGGAGCCCGTCTTGCTGCCCAAGGAG   503

Qy 1942 AlaProArgValAlaArgProGluArgProAlaAspThrGlyArgAlaAspPheHisLeuAla 1961
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   502 GCCCCCCGGGTCGCCCGGCCAGAGCGGCCCGAGCAGACACCGGCCATGCCTTCCTGCC   443

Qy 1962 LysProProAlaArgSerGlyLeuGluProAlaSerSerProSerGlySerGlyGluPro 1981
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   442 AAGCCCCCAGCCCGCTCCGGGCTGGAGCCCGCCTCCTCCCCAGCAAGGGCTCGGAGCCC   383

Qy 1982 ArgProLeuValLeuProProValSerGlyHisAlaThrIleAlaAlaArgThrProAlaLeuAsn 2001
          ||||||:::||| |||||||||||||||||||||||||||||||||||||||||
Db   382 CGGCCCCATAGTGCCTCCTGTCTCGGCCCACGCACCCCTGCGAAGAAC   323

Qy 2002 LeuAlaProHisHisAlaSerProAspProProAlaProProAlaSerAlaSerAspPro 2021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 CTCGCACCTCACCACGCCAGCCCAGCCCGGACCCCGGCCGCCCTGCCTCGGACCCG   263

Qy 2022 HisArgGluLysThrGlnSerLysProPheSerIleGlnLeuGluLeuArgSerLeu 2041
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262 CACCGGGAAAAGACTCAAAGTAAACCCTTTCCATCCAGAACCTGGAACTCCGTTCTCTG 203

Qy 2042 GlyTyrHisGlySerSerTyrSerProGluGlyValGlyProValSerProValSerSer 2061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   202 GGTTACCACGGCAGCAGCTACAGCCCCGAAGGGGTGGAGCCCGTCAGCCCCGTGAGCTCA 143

Qy 2062 ProSerLeuThrHisAspLysGlyLeuProLysHisHisLeuGluLeuGluAspLysSerHis 2081
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142 CCCAGTCTGACCCACGACAAGGGGCTCCCCAAGCACCTGGAAGAGCTGACAAGAGCCAC   83

Qy 2082 LeuGluGlyGlyLeuLysArgProLysGlnProGlyProValLysLeuGlyGlyGluAlaAla 2101
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 CTGGAGGGGGGAGCTGCGGCCCAAGCAGCCAGCCCCCCGTGAAGCTTGGCGGGGAGGCCGCC   23

Qy 2102 HisLeuProHisGluArgPro 2108
          |||||||||||||||||||||
Db   22 CACCTCCCACACCTGCGCGCCG   2

RESULT 82
BU108220
LOCUS   BU108220   1088 bp   mRNA   linear   EST 25-NOV-2002
DEFINITION 603112484F1 CSEQCHL12 Gallus gallus cDNA clone ChEST61g22 5', mRNA
            sequence.
ACCESSION   BU108220
VERSION   BU108220.1 GI:25311097
KEYWORDS   EST.
SOURCE   Gallus gallus (chicken)
  ORGANISM Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 1088)

```
Seq primer: -40UP from Gibco
High quality sequence stop: 498.
                               1..592
                    Location/Qualifiers
FEATURES
     source          1..592
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3645044"
                     /tissue_type="fibrorheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Ov18"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCCGCCGACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.:                    592
Score:                        1.49e-34
Percent Similarity:           1053.00          Matches:          197
Best Local Similarity:        100.00%          Conservative:     0
Query Match:                  100.00%          Mismatches:       0
DB:                           7.97%            Indels:           0
                              10               Gaps:             0

US-09-522-753-5 (1-2517) x BF436351 (1-592)
```

```
Qy   2113 GlnProSerSerProLeuLeuGlnThrAlaProGlyValLysGlyHisGlnArgVal 2132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 CAGCCGTCGTCCAGCCCGCTGCTCCAGACCGCCCCAGGGGTCAAAGGTCACCAGCGGGTG 61

Qy   2133 ValThrLeuAlaGlnHisHisSerGluValIleThrGlnAspTyrThrArgHisHisPro 2152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 GTCACCCTGGCCCAGCACCATCAGTGAGGTCATCACACAGGACTACACCCGGCACCACCCA 121

Qy   2153 GlnGlnLeuSerAlaProLeuProAlaProLeuTyrSerPhePheProGlyAlaSerCysPro 2172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 CAGCAGCTCAGCGCACCCCTGCCCGCCCCCCTCTACTCCTTCCCTGGGGCCAGCTGCCCC 181

Qy   2173 ValLeuAspLeuArgArgProProSerAspLeuTyrLeuProProAspHisGlyAla 2192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 GTCCTGGACCTCCGCCGCCCACCCAGTGACCTTCTACCTTCCCGCCCGACCATGGTGCC 241

Qy   2193 ProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsnLysThr 2212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CCGGCCCGGGGCTCCCCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAAGACG 301

Qy   2213 SerValLeuGlyGlyGlyValGlyAspGlyIleLeuGluProValSerProGluProGlyMetThr 2232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGGCATGACG 361

Qy   2233 GluProGlyHisSerArgSerArgAlaValAlaTyrProLeuLeuTyrArgAspGlyGluGlnThr 2252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 GAGCCAGGGGCACTCCCGGAGTGTGTTGCTTACCCGCTGCTCTGTACCGGGGATGGGAACAGACG 421

Qy   2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnInProProAlaPhePhe 2272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 GAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACACCAGCCGCCGCAGCCTTCTTC 481

Qy   2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysLysGlnGluIleAsnLys 2292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 AGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAAGCAAGAGATCAACAAG 541

Qy   2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnPro 2309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 AAGTTGAACACCCACAACCCGAATGAGCCTGAATACAATATCAGCCAGCCT 592

RESULT 84
```

```
Qy   1256 AspArgGlyArgGluAspSerLeuProLysGlyHisValIleTyrGluGlyLysLysGly 1275
          :::::: ||||||||||:::::::||||||||||| ||| ||||||||
Db    421 GAGAAAGTGAGAGGAGAGATGCCATGCCCAAAGGACCATGTCATTTACGAGGGGACGAAAGGC 480

Qy   1276 HisValLeuSerTyrGluGlyMetSerValThrGlnCysSerLysGluAspGlyArg 1295
          |||||||||:: |||||||||||| :::|||| ||||||||||||||||||||||||
Db    481 CACGTGCTGTCCTACGACGGTGGGGTTGCTGTCTTCAGTGTCCCAAAGAGAACAGCAGG 540

Qy   1296 SerSerSerGlyProProHisGluThrAlaAlaProLysArgThrTyrAspMetMetGlu 1315
          ||||||  ||||||
Db    541 ---AGCTCAGGAGCTTCCATGAGCCCATGCCACGGAACCAAGCGACCTACGATATGATGGAG 597

Qy   1316 GlyArgValGlyArgAlaIleSer----------SerAlaSerIleGluGlyLeu 1330
          ||||:: ||| :::|||| ||||||||||||||
Db    598 GGGAGGATCAGCCGTGGCTTGTCATCCCGGGACCCGCTGTCGCCAGCATTGAAGGTCTC 657

Qy   1331 MetGlyArgAlaIleProProGluArgHisSerProHisLeuLysHisGluGlnHisHis 1350
          |||||| || ||||||||:::||:::||||
Db    658 ATGGGTCGAGCCATCCCTCCAGAACCAGACACAGTCCCATAACCTGAAGGAGCAGCACCAC 717

Qy   1351 IleArgGlySerIleThrGlnIleIleProArg-SerTyrVal-GluAlaGlnGlu--Asp 1369
          |||||||||||||||||||| ||| || || |||
Db    718 ATCCGGGGCTCCATCACACAAGAATACCTCGGGTCCTATGTCGGAGGCCCATGAGTGAC 777

Qy   1370 TyrLeuArgArgGluAlaLysHisLeu-LysArgGlyThrProProProProProPr 1389
          ||||:: ||||:::|||| ||||| ||||| :::||||||
Db    778 TACCTCAGAAGAGAAGCAAACAGCTCCAAGAGAGGGAAAACACTTCCACCA---------- 826

Qy   1389 oSerArgAspLeuThrGluAlaTyrArgLysThrGlnAlaLeu------------GlyProLeu----- 1405
          :::::||||||| ||:: |||
Db    827 ----CGGGACTTATCTGACGCATACAAGGT-CACGGTCTCACCGAGGGGCCCTTACTCCC 881

Qy   1406 ------LysLeuLysProAlaHisGlyGly----LeuValAlaThrValLysGluAlaGl 1422
          ::: ||||||||
Db    882 CCTGAAAAATTGAAGGCCAGGCCACGATGGGCCTTGGTTGGCCACCGGTGAAAGGAA-GC 940

Qy   1422 yArgSerIleHisGluIle--------ProArgGluGluLeuArgHis---ThrProGl 1438
          ||| :::
Db    941 ACGGAAGGTTCAACCCATTGAGAATCCCCGAAGAAGGAGCTGAGGCGGGACACCGGCGAA 1000

Qy   1438 uLeuProLeuAlaProArgProLeuLysGlySerIleThrGlnGlyThrPro 1456
          ::: ||||||||
Db   1001 CATCTCTCTGGCGAAGGGCCTCTTGAAGGAGGAAGGTCCCACTCACCCAGGGTCCCCCT 1055

RESULT 83
```

```
BF436351
LOCUS           BF436351                592 bp    mRNA    linear   EST 29-NOV-2000
DEFINITION      7p06e11.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3645044 3'
                similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
                , mRNA sequence.
ACCESSION       BF436351
VERSION         BF436351.1   GI:11448666
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 592)
  AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
  JOURNAL       Unpublished (1997)
  COMMENT       Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
                R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
                Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
                I.M.A.G.E. Consortium DNA Sequencing by: Washington University
                Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL, send email to:
                info@image.llnl.gov
```

```
BI735071
LOCUS       BI735071                 820 bp    mRNA     linear   EST 20-SEP-2001
DEFINITION  603356210F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363282 5',
            mRNA sequence.
ACCESSION   BI735071
VERSION     BI735071.1  GI:15712084
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11924 row: d  column: 03
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..820
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5363282"
                     /tissue type="retina"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_94"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 3.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:                                       Length:      820
Score:              7.74e-34                      Matches:     213
Percent Similarity: 1039.50                       Conservative: 8
Best Local Similarity: 87.35%                     Mismatches:  25
Query Match:        84.19%                         Indels:      9
DB:                 7.87%                          Gaps:        3
                    12

US-09-522-753-5 (1-2517) x BI735071 (1-820)
```

```
Qy  1929 TyrProThrLeuMetGluProLysGluLeuProValLeuLeuProLysGlyGlySerGluProAlaProArgValAlaArgPro 1948
             ||||||||||||||||||| |||  ||||||||||||  |||   |||||||||||||||||||
Db     1 TACCCTACCCTCATGGAGCCCCGTCTTACCCAAGGAGACCTTCGGGTCGCCCGGCCC                                60

Qy  1949 GluArgProArgAlaAspThrGlyHisGlyHisAlaPheLeuAlaLysProProAlaArgSerGly 1968
             |||||||||| ||| |||||||||||||||  |||||||||||||||||||||||
Db    61 GAGCGGCCCCGTGGACGCGGCTGGCCATGCCTTCCTCACCAAACCCCGGCCCGG------         114

Qy  1969 LeuGluProAlaSerProSerLysGlySerGluProArgProLeuValProProVal 1988
             |||||||||||||||||||  ||||  ||||||||| ||| |||||||||||
Db   115 --GAGCCCGCCTCCTCACCCAGGCAAGAGCTCCGAGCCCGATCCCTTAGCACCCCCAGC 171

Qy  1989 SerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaProHisHisAlaSer 2008
             |||||||| ||||||||||||||||||||||||  |||||||||||
Db   172 TCCAGGCACACAGCCATCGCCCGCACCCCAGCAAAGAACCTTGCACCCCACCATGCCAGT 231

Qy  2009 ProAspProProAlaProProAlaSerAlaSerAspProHisArgGluLysThrGlnSer 2028
             |||||||| ||| |||||||||||||||||||||||||||||||
Db   232 CCGGACCCCGCGGCG---CCCACCTCGGCCTCAGATCTGCACCGAGAAAAGACTCAAAGT 288

Qy  2029 LysProPheSerIleGlnGluGluLeuGluLeuArgSerLeuGlyLeuGlyTyrHis--GlySerSer 2047
```

```
Db   289 AAACCCTTTTCCATCCAGGAATTGAACTCCGTTCTCTGGTTCCACAGTGGAGCTGGC 348
             ||||||||||||||||||||||||||||||||||||||||||||:::
Qy  2048 TyrSerProGluValGluProValSerProValSerProSerLeuThrHisAsp 2067
             ||||||||:::|||||||||::::|||||||||:|||::||||||
Db   349 TACAGCCCCGATGGGTGGAGCCCATCAGCCCGGTGAGCTCCCCAGCCTGACCCACGAC 408
Qy  2068 LysGlyLeuProLysHisLeuGluGluLeuAspLysSerHisLeuGluGlyGluLeuArg 2087
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   409 AAGGGGCTCCCAAACCTGGAAGAGCTAGAGAAGAGCCACTTGGAAGGGGAGCTGCGG 468
Qy  2088 ProLysGlnProGlyProValLysProValLysGluLysAlaLeuHisLeuProHisLeuLys 2107
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   469 CACAGCAGCCAGCCCCCATGAAGCTCAGCGCGGAGGCTGCCCATCTCCCACATCTGCGG 528
Qy  2108 ProLeuProGluSerProSerSerProLeuLeuGlnThrAlaProGlyValLys 2127
             |||||||||||||||||||||||||||||||||||||||||||||||::|||
Db   529 CCACTGCCCGAGAGCCCAGCCCTCCATCTAGCCCACTCCTCCAGACTGCCCCAGGCATCAA 588
Qy  2128 GlyHisGlnArgValValThrLeuAlaGlnHisHisLeuSerGluValValIleThrGlnAspTyr 2147
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   589 GGTCACCAGAGGGTGGTCACCTTGGCTCAGCACATCAGTGAGGTCATTACGCAGGACTAC 648
Qy  2148 ThrArgHisHisProGlnGlnLeuSer--AlaProLeuProAlaProLeuTyrSerPheP 2167
             |||||||||||||||||||||::||||||||||||||||
Db   649 ACGGCGCACGA-CCGCAGCAGCCCAAGTGCCCCCCTCCCCGCCCCTCTTCACTCCTTTC 707
Qy  2167 roGlyAlaSerCysProValLeuAspLeuArgArg 2178
             |||||||||||::|||||||||:::
Db   708 CCGGAGCAGCTGG-CCTGTCCGGGATCTTCGGCAG 741
```

```
RESULT 85
AW438580
LOCUS       AW438580                 611 bp    mRNA     linear   EST 14-FEB-2000
DEFINITION  xs82e10.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2776170 3'
            similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
            ;, mRNA sequence.
ACCESSION   AW438580
VERSION     AW438580.1  GI:6973886
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Seq primer: -40UP from Gibco
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1..611
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2776170"
                     /tissue type="moderately-differentiated endometrial
                     adenocarcinoma, 3 pooled tumors"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Ut2"
```

```
              /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
              Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
              Average insert size 1.85 kb. Life Technologies catalog #:
              11539-012"

ORIGIN

Alignment Scores:
Pred. No.:               6.46e-34          Length:          611
Score:                   1038.00          Matches:          194
Percent Similarity:      96.06%           Conservative:     1
Best Local Similarity:   95.57%           Mismatches:       8
Query Match:             7.85%            Indels:           0
DB:                                       Gaps:             0

US-09-522-753-5 (1-2517) x AW438580 (1-611)

Qy  2118 ProLeuGlnThrAlaProGlyHisGlnArgValValThrLeuAlaGln 2137
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db     3 CCGCTGCTCCAGACCGCCCCAGGGTCAAAGGTCACCAGCGGGTGGTCACCCTGGCCCAG 62

Qy  2138 HisIleSerGluValIleThrArgGlnAspTyrThrArgHisHisProGlnGlnLeuSerAla 2157
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    63 CACATCAGTGAGGTCATCACACAGGACTACACCCGGCACCACCCACCAGCAGCTCAGCGCA 122

Qy  2158 ProLeuProAlaProLeuTyrSerPheProGlyAlaSerCysProValLeuAspLeuArg 2177
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   123 CCCCTGCCGCCCCCCCTCTACTCCTTCCCTGGGCCAGCTGCCCCGTCCTGGACCTTCCGC 182

Qy  2178 ArgProProSerAspLeuTyrLeuProProProAspHisGlyAlaProAlaArgGlySer 2197
         ||||||||||||||||||||||||||||||||||||||||||||
Db   183 CGCCCACCCAGTGACCCTTACCTCCCCGCCCCCCGACCCGTGCCCCCCGTGGCTCC 242

Qy  2198 ProHisSerGluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGly 2217
         |||||||||||||||||||||||||||||||||||||||||
Db   243 CCCCACAGCGAAGGGGGCAAGAGGTCTCCAGAGCCAAACAACACGTCGGTCTTGGGTGGT 302

Qy  2218 GlyGluAspGlyLysGluGluProValSerProGluProProGlyMetThrGlyProGlyHisSer 2237
         |||||||||||||||||||||||||||||||||||||||||||
Db   303 GGTGAGGACGGTATTGAACCTGTGTCCCCACCGGAGGGCCATGACGAGCCCCAGGGCCACTCC 362

Qy  2238 ArgSerAlaValAlaTyrProLeuLeuTyrArgAspGlyGlnThrGluGlnThrGluProSerArgMet 2257
         |||||||||||||||||||||||||||||||||||||
Db   363 CGGAGTGCTGTGTCATACCCGCTGCTGTACCGGGATGGGCAGACAGAGCAGACAGGGATG 422

Qy  2258 GlySerLysSerProGlyAsnThrSerGlnProGlyProAlaPhePheSerLeuProThrGlu 2277
         ||||||||||||||||||||||||||||||||||||||
Db   423 GGCTCCAAGTCTCCAGGCAACACCAGCCAGCCCGGCCCAGCCTTCTTCAGCAAGCTGACCGAG 482

Qy  2278 SerAsnSerAlaMetValLysSerLysGlnIleLeuAsnLysLysLeuAsnThrHis 2297
         ::::|||||||||||||||||||||||||||||||||||||
Db   483 AAGCACTCCGTCAGTGCCATGGTGAAGAGCAAGAAGATCAACAAGAAGCTGAACACTCAC 542

Qy  2298 AsnArgAsnGluProGluGlnIleSerGlnProGlyGlnProThrGlyIleProGluPheAsnMetPro 2317
         |||||||||||||||||||||||||||||||||||||||
Db   543 AACCGGAATGAGCCTGAATACAATATTCGCCAGCCTGGACGGAGAAATCTTAATATGCCC 602

Qy  2318 AlaIleThr 2320
         |||||||||
Db   603 GGCCATCACC 611

RESULT 86
BG391632
LOCUS         BG391632   602417893F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4537438 5',
DEFINITION    624 bp    mRNA    linear   EST 12-MAR-2001
              mRNA sequence.
ACCESSION     BG391632
VERSION       BG391632.1   GI:13285080
KEYWORDS      EST.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 624)
```

```
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10462 row: i column: 23
           High quality sequence stop: 623.

FEATURES            Location/Qualifiers
     source         1..624
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4537438"
                    /tissue_type="embryonal carcinoma, cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH MGC 92"
                    /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                    Average insert size 2.5 kb. Library enriched for
                    full-length clones and constructed by Life Technologies.
                    Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:               6.61e-34          Length:          624
Score:                   1038.00          Matches:          206
Percent Similarity:      99.04%           Conservative:     0
Best Local Similarity:   99.04%           Mismatches:       2
Query Match:             7.85%            Indels:           2
DB:                                       Gaps:             0

US-09-522-753-5 (1-2517) x BG391632 (1-624)

Qy  1221 SerAlaIleThrTyrArgGlySerIleThrHisGlyThrProAlaAspValValLeuTyrLys 1240
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3 AGCGCCATCACATACCGCGGCTCCATCACCCACGGCACGCCAGCTGACGTCCTGTACAAG 62

Qy  1241 GlyThrIleThrArgIleIleGluGlySerProSerArgLeuAspArgLeuAspArgGlyArgGlu 1260
         |||||||||||||||||||||||||||||||||||||||||||||||
Db    63 GGCACCATCACCAGGATCATCGGCGAGAGCACAGCAGCGCCGTGGACCGGGACCGGGGAG 122

Qy  1261 AspSerLeuProLysGlyHisValIleTyrGluLysLysGlyLysHisValLeuSerTyr 1280
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   123 GACAGCCTGCCCAAGGGCCACGTCATCTACGAGAAGAAGGGCCACGTCTTCTGTCCTAT 182

Qy  1281 GluGluGlyMetSerValThrGlnCysSerLysSerGluSerGlyLysAspGlyAspGlyArgGlyValGlyArg 1300
         |||||||||||||||||||||||||||||||||||||||||
Db   183 GAGGGTGGCATGTCTGTGACCCAGTGCTCCAAGGAGGAGCGGCAGAAGCAGCTCAGGACCC 242

Qy  1301 ProHisGluThrAlaAlaProLysArgThrTyrAspMetGluGlyValGlyArg 1320
         ||||||||||||||||||||||||||||||||||||||||||||
Db   243 CCCATGAGACCGCCGCCCCAAGCGCACCTATGACATGATGAGGGGCCGTGGGCAGA 302

Qy  1321 AlaIleSerSerAlaSerIleGluGluLysMetGlyArgAlaIleProProGluArgHis 1340
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 GCCATCTCCTCAGCCAGCATGGAAGGTCTCATGGGCCGTGCCATCCCGCCGGAGCGACAC 362

Qy  1341 SerProHisGlyLeuLysGlyGlnHisHisIleArgGlySerIleThrGlnIleGlyIlePro 1360
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   363 AGCCCCCACCACCTCAAAGAGCAGCACCACCATCCGCGGGTCCATCACACAAGGATCCCT 422

Qy  1361 ArgSerTyrValGlnAlaGlnAlaGlnAspTyrLeuArgArgGluAlaLysLeuLysArg 1380
         |||||||||||||||||||||||||||||||||||||||||||||
Db   423 CGGTCCTACGTGGAGGCACAGGAGCACAGGAGGCCAAGCTCCTGAGGCCAAGCTCCTAAAGCGG 482

Qy  1381 GluGlyThrProProProProSerProArgArgAspLeuThrGluAlaTyrArgLysThrGln 1400
```

US-09-522-753-5 (1-2517) x CB518985 (1-668)

```
Qy   1418 ValLysGluAlaGlyAlaHisGlyArgSerIleHisGluIleProArgGluGluArgHisThrPro 1437
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTGAAGGAGGCGGGGCGCGTCTATCCATGAGATCCCGAGAGAGAGCTGCGCCGCCACCN   60

Qy   1438 -GluLeuProLeuAlaProArgProLeuLysGluGlyLySerIleThrGlnGlyThrProLe 1457
          |||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGAGCTACCCTGGCCACCACGGCCTCTGAAGGAGGGTTCCATCACCCAGGGCACCCCACT  120

Qy   1457 uLysTyrAspThrGlyAlaSerThrGlyLySerLysHisAspAlaValArgSerLeuIl 1477
          |||||||||||||||||||  :::|||||||||||||||||||||
Db    121 CAAGTACGACTCTGGGGCACCCTCACTGGCACCAGAAACACGACGTGCGCTCCATCAT  180

Qy   1477 eGlySerProGlyArgThrPheProProValHisProLeuAspValMetAlaAspAlaAr 1497
          |||||||||||   ||||| :::|||||||||||||||||||||||||
Db    181 CGGCAGCCCCGGCCGCCCTTCCCTGCCCGTGCACCCGCTGGACATAATGGCTGACGCCG  240

Qy   1497 gAlaLeuGluArgAlaCysTyrGlyGluSerLeuLySerArgProGlyThrAlaSerSe 1517
          ||||||||||||||||||||||||||||    |||||||||||||||||
Db    241 GGCACTGGAGCGTGCCTGCTACGGGGAGTCTGAAGAGCCGGTCAGGGACCAGCAGTGG  300

Qy   1517 rSerGlyGlySerIleAlaArgGlyAlaProValIleValProGlyLeuGlyLysProAr 1537
          |||||||||||||||||||||:::|||||||||||||||||||||
Db    301 TGCAGGGGGCTCCATCACACGTGGGGCTCCAGTCGTCGTGCTGGCAAGCCACG  360

Qy   1537 gGlnSerProLeuThrTyrGluAspHisGlyAlaProPheAlaGlyHisLeuProArgGl 1557
          ||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCAAAGCCCACTGACTTACGAAGACCACGGGGCACCCCTTCACCAGGTCACTGCCACGTGG  420

Qy   1557 yYSerProValThrMetArgGluProThrProArgLeuGlnGlyGlySerLeuSerSerSe 1577
          ||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTCCCGTGTGACCACGAGGAGCCCACGCCTTCAGGAAGCGAGCCTCCTATCCAG  480

Qy   1577 rLysAlaSerGlnAspArgArgLeuThrSerThrProArg-GluIleAlaAlaLySerProH 1597
          ||||||||||||||||||||||||||  |||||||||||
Db    481 CAAGGCTAGCCAGGACCCGGAAGCTGACATCTACACCCCGGNAGATCGCCAAGTCCCCAC  540

Qy   1597 isSerThrValProGluHisHisProIleSerProTyrGluHisProLeuLeuArgG 1617
          ||||||||||||||||||||||||||||||||||||||||
Db    541 ACAGCACTGTGCCCGAGCACCACCCTCACCCCATCTCCCCCTATGAGCACTTGCTCCGGN  600

Qy   1617 lyValSerGlyValAspLeuTyrArgSerHisIleProLeuAlaPheAspProThrSerI 1637
          |||:::|||||||||||||||||||||||||||||||||||
Db    601 GCGTGACTGGTGTGGACCTGTACCGGTCACATCCCATTGGCCTTTGACCCCACCTCCA  660

Qy   1637 leProArg 1639
          ||||||
Db    661 TACCCCGA 668
```

```
Db    483 GAGGGCAGCCGCTCCGCCCACCGGCCCCTCCACGGGACCTGACGLuGlyLeuVAlAlaThrValLysGlu 542
          |||||||||||||||||||||||||||||||||||||||||||||||

Qy   1401 AlaLeuGlyProLeuLysGluLysLeuGlyArgProAlaHisGlyGluGlyLeuVAlAlaThrValLysGlu 1420
          ||||||||||||||||||||||||   |||||||||||||||||||
Db    543 GCCCTGGGCCCCCTGAAGGTGAAGCCGGC-CATGAGGGGCTGGTGGCCACGGTGAAGGAG  601

Qy   1421 AlaGlyArgSerIleHisGluIle 1428
          ||||||||||||  |||||||
Db    602 GCGGGCCGTCC-ATCCATGAGATC  624
```

RESULT 87
CB518985

LOCUS   CB518985   668 bp   mRNA   linear   EST 09-JUL-2003
DEFINITION UI-M-GH0-cee-f-04-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
            IMAGE:6838373 5', mRNA sequence.
ACCESSION   CB518985
VERSION   CB518985.1 GI:29352340
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 668)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousefl.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES   Location/Qualifiers
     source   1..668
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6838373"
                     /tissue_type="Whole brain"
                     /dev_stage="1, 5, and 15 days newborn"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_GH0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is CGAACTGAAT. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   7.12e-34
Score: 1038.00
Percent Similarity: 93.30%   Length: 668
Best Local Similarity: 88.84%   Matches: 199
Query Match: 7.85%   Conservative: 10
DB: 14   Mismatches: 13
                                               Indels: 2
                                               Gaps: 0

RESULT 88
CD244013

LOCUS   CD244013   925 bp   mRNA   linear   EST 22-MAY-2003
DEFINITION AGENCOURT_14096335 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30377707 5', mRNA sequence.
ACCESSION   CD244013
VERSION   CD244013.1 GI:31004477
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 925)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
```

```
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: NDAM442  row: f  column: 20
        High quality sequence start: 21
        High quality sequence stop: 560.

FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3037707"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
                     Site 2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."

ORIGIN

Alignment Scores:
Pred. No.:          1.12e-33         Length:          925
Score:              1037.00          Matches:         239
Percent Similarity: 64.34%           Conservative:    10
Best Local Similarity: 61.76%        Mismatches:      31
Query Match:        7.85%            Indels:          108
DB:                 14               Gaps:            9

US-09-522-753-5 (1-2517) x CD244013 (1-925)

Qy   2146 AspTyrThrArgHisHisProGlnLeuSerAlaPro---LeuProAlaProLeuTyr 2164
          :::::     ||  ::    :::   ||||  :::       |||  |||  ||
Db     18 GAATTCTTGCGGATGCATCATCGACCGTCCAAGTGCCACACCTGCCTGTGCTGCGTGCCC 77

Qy   2165 SerPheProGlyAla-------------------------------SerCysPro 2172
          ||||||||||                                       |||||||
Db     78 CCGACACCAGGCACCCCAGACCACCCGCATCCACGACCGGCCTTGCCTACCTCCCCAGCTGCCCC 137

Qy   2173 ValLeuAspArgArgProSerAspLeuTyrLeuProProProAspHisGlyAla 2192
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    138 GTCCTGGACCTCCGCCGCCCCAGTGACCTCTACCTCCCGCCCCCCGGACCATGGTGCC 197

Qy   2193 ProAlaArgGlySerProHisSerGluGluGlyGlyLysArgSerProGluProGlnLysThr 2212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    198 CCGGCCCGTGGCTCCCCCAGCGAAGGGGCAAGAGGTCTCCAGAGCCAAACAAGACG 257

Qy   2213 SerValLeuGlyGlyGlyValGluAspGlyIleGluGluProValSerProProGlyMetThr 2232
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    258 TCGGTCTTGGGTGGTGGTGAGGACGGTATTGAACCTGTGTCCCCACCCGGGCCATGACG 317

Qy   2233 GluProGlyHisSerArgSerArgAlaValTyrProLeuLeuTyrArgAspArgArgGlyGlnThr 2252
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    318 GAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGAACAGACG 377

Qy   2253 GluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAlaPhePhe 2272
          ||||||||||||||||||||||||||||||||||||||||||||
Db    378 GAGCCCAGCAGGATGGGCTCCAAGTCTCCAGGCAACACACCAGCCGCGCCAGCCTTCTTC 437

Qy   2273 SerLysLeuThrGluSerAsnSerAlaMetValLysSerLysGlnIleLeuAsnLys 2292
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    438 AGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAGAAGCAAGATCAACAAG 497

Qy   2293 LysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIleSerGlnProGlyThrGlu 2312
          ||||||||||||||||||||||||||||||||||||||||||||
Db    498 AAGCTGAACACCACAAACCGAAATGAGCCTGAATACAATATCAGCCAGCCTGGGACGGAG 557

Qy   2313 IlePheAsnMetProAlaIleThrGlyLeuMetThrTyrArgSerArgSerGlnAlaVal 2332
          ||||||||||||||||||||||||
Db    558 ATCTTCAATATGCCCGCCATCACCGGAACA------------------------------ 587

Qy   2333 GlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeuMetGly 2352
```

```
Db    587 ---------------------------------------------------------- 587

Qy   2353 LysTyrAspGlnTrpGluGluSerProProLeuSerAlaAsnAlaPheAsnProLeuAsn 2372

Db    587 ---------------------------------------------------------- 587

Qy   2373 AlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgSerArgAspHis 2392

Db    587 ---------------------------------------------------------- 587

Qy   2393 ThrLeuThrSerProGlyGlyGlyLysSerAlaLysSerGlyArgProSerArg 2412
                                               ||||||||||||||
Db    588 ----------------------GGTGGCGGCGGAAGGGCAAGGTCTGGCAGACCAGC-AGCCGA 631

Qy   2413 LysAlaLysSerProAlaPro-GlyLeuAlaSerGlyAspArgPro-ProSerValSerS 2432
          ||||||||||||||||||    ||||||||||||||||||||||   |||||||||
Db    632 AAAGCCAAGTCCCGGCCCCCGGGGCTGGCATCTGGGGACCGGCCACCCCTCTGTCTCCT 691

Qy   2432 erValHisSerGlu-GlyAspCysAsnArgArg-ThrProLeu-ThrAsnArgValTrpG 2451
          :::||:::::| ::|     ||    :::|     :::    ||:|||
Db    692 CAGTGCACTCGAAGGGAGACTGCAACCGCCGGAGACGCGCCTCAACAAACCGGTGTGGG 751

Qy   2451 luAspArg---ProSerSer-AlaGlySerThrProPhePro---TyrAsnProLeuIle 2468
          ::   ::|     |||||     ||||||     ||     ||
Db    752 AAAGAAAGGGCGCCTCGTTCCCGCAGGGTTCCCGGCCATTCCCTTACAAACCCCTGAATC 811

Qy   2469 MetArgLeuGlnAlaGlyValMet------AlaSerProProProGly---LeuPro 2485
          ||||||||||||||||::::|          ||::::||||||||     |||||
Db    812 ATGCGGGTTGCAGGCCGGGTGTCAATGGCCGTCCCCACCCCCCCGGGGGCCCTTCCC 871

Qy   2486 AlaGly------SerGlyProLeuAlaGlyProHisHisAlaTrpAsp-GluGluProLy 2503
          ||||||          :::     ||||||         ::::|
Db    872 CCCGGGGCAGCCGGCGGGGGCCCCTT------------------CGCTGGGGCCCCCAACCAAG 916

Qy   2503 sProLeu 2505
          :||||||
Db    917 GCCCTTG 923
```

```
RESULT 89
BG119261
LOCUS           BG119261    1030 bp    mRNA    linear    EST 30-JAN-2001
DEFINITION      602349134F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443969 5',
                mRNA sequence.
ACCESSION       BG119261
VERSION         BG119261.1  GI:12612767
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 1030)
  AUTHORS       NIH-MGC http://mgc.nci.nih.gov/,
                National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE         Unpublished (1999)
  JOURNAL
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: ATCC
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM10219  row: c  column: 10
                High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1..1030
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4443969"
                     /tissue_type="adenocarcinoma, cell line"
```

```
LOCUS       AW701437    654 bp    mRNA    linear    EST 18-APR-2000
DEFINITION  uq86b05.y1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:2937969 5'
            similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
            ; mRNA sequence.
ACCESSION   AW701437
VERSION     AW701437.1  GI:7585568
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 654)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            MGI:1050389
            Seq primer: -40RP from Gibco
            High quality sequence stop: 470.
FEATURES             Location/Qualifiers
     source          1..654
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="Czech II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2937969"
                     /tissue_type="pooled lung tumors"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NCI_CGAP_Lu33"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TCTTACCAATCTGAAGTGGGAGCGGCGCCCCTCGTTTTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
```

ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.07e-33
Percent Similarity: 1033.50   Length: 654
Best Local Similarity: 78.49%   Matches: 205
Query Match: 77.36%   Conservative: 3
DB: Mismatches: 10
                                            Indels: 47
                                            Gaps: 2

US-09-522-753-5 (1-2517) x AW701437 (1-654)

```
Qy  2188  ProAspHisGlyAlaProAlaArgGlyProHisSerGluGlyGlyLysArgSerPro  2207
          ||||     |||  ||   |  |||   ||||||::: |||||||||  |||||
Db     1  CCCGACCATGGCTCCCCAGCCCGGGATCCCCCACAGTGAAGGGGCAAAAGGTCCCCA    60

Qy  2208  GluProAsnLysThrSerValLeuGlyLysGlyGlyIleGluProValSerPro  2227
          ||||:::|||   ||| |||||  ||||||||  ||||  ||||||  |||||
Db    61  GAACCCAGCAAAACATCGGTCCTGGGCAGTGGGATGCCATTGAGCCTGTGTCCCCA   120

Qy  2228  ProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuTyrArg  2247
          ||||||||  |||:: ||||||  |||||:: |||  ||:: |||||||  |||||
Db   121  CCAGAGGGCATGACTGAGCCAGGACATGCTCGGAGCGCGCTGGTGTACCCACTGCTGTATCGA   180
```

                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_90"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:
Score: 1.53e-33
Percent Similarity: 1035.00   Length: 1030
Best Local Similarity: 86.38%   Matches: 213
Query Match: 82.88%   Conservative: 9
DB: Mismatches: 32
                                            Indels: 5
                                            Gaps: 2

US-09-522-753-5 (1-2517) x BG119261 (1-1030)

```
Qy  2201  GluGlyGlyLysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyGluAsp  2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  GAAGGGGCAAGAGGTCTCCAGAGCCAAACAAGACGTCGGTCTTGGGTGGCGGTGAGGAC    60

Qy  2221  GlyIleGluProValSerProGluProGlyMetThrGluProGlyHisSerArgSerAla  2240
          |||||||||||||||||||||||||||||||  ||||||||| |||||||||
Db    61  GGTATTGAACCTGTCGTCCCACCGGAGGGCATGACGGAGCCAGG-CACTCCCGGAGTGCT   119

Qy  2241  ValTyrProLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLys  2260
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   120  GTGTACCCGCTGCTGTACCGGGATGGGGAACAGACGGAGCCCAGCAGGATGGGCTCCAAG   179

Qy  2261  SerProGlyAsnThrSerGlnProProAlaPhePheSerLysLeuThrGluSerAsnSer  2280
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   180  TCTCCAGGCAACACCAGCCAGCCGCCAGCCTTCTTCAGCAAGCTGACCGAGAGCAACTCC   239

Qy  2281  AlaMetValLysSerLysGlyGlnIleGluLysAsnLysLysAsnThrHisAsnArgAsn  2300
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   240  GCCATGGTCAAGTCAAGGGAACAAGAGATCAACAAGAAGTCAACACCCACACCGGAAT   299

Qy  2301  GluProGluGlyThrGlyIlePheSerGlnProGlyThrGlyIlePheAsnMetProAlaIleThr  2320
          |||||||   |||  |||  ||||||  ||||||:: |||||||||||
Db   300  GAGCCTGAATACAACATATCGCAGCCTGGGACGGAGATCTTCAATATGCCCGCCATCACC   359

Qy  2321  GlyThrArgLeuMetThrTyrArgSerGlnAlaValGlnGluHisAlaSerThrAsn-Me  2340
          |||  ||||||||||||||||||  |||  ||||||||||||  |||:: |||
Db   360  GGAACAGAGCCTTATGACCTATAGAAGCCAGGCGGTGCAGGAACATGCCAGAAACCAACATT   419

Qy  2340  tGlyLeuGluAlaIleIleArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluGluSe  2360
          :|||||||||||||||||||||  ||||||:: |||||||||||||||||||||||
Db   420  GGGGCTGGAGGCCATAATTAGAAAGGCACTCATGGGTAAATATGACCAGTGGGAAGAGTC   479

Qy  2360  rProProLeuSerAlaAlaAsnAlaPheAsnProLeuAsnAlaSerAlaSerLeuProAlaAl  2380
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   480  -CCGCCGCTTCAGCGCCAATGCTTTTAAACCTCTGAATGCCAGTGCCAGACTGCCGTGCT   538

Qy  2380  aMetProIleThrAlaAlaAspThrAlaSerAspHisThrLeuThrSerProGlyGlyGl  2400
          |||||||||||||||||:::|||  :::||   ||:|||||||||||||||  |||||
Db   539  ATGCCAATTACCGCTGCTTGACGGAGTGACACTTCCCCAGCGTGGCGG   598

Qy  2400  yGlyLysAlaLysValSerGlyTyrArgProSerSerArgGlyArgAlaLysSerProAlaProGl  2420
          ||||||||||||||||||||   :::||||||  |||||||  ||||||||||||||||||
Db   599  CGGGAAGGCCAAGGTCTCTGGACAGCCAGAGAGG--AAAAGCCAAGTCCCGGCCCCGGGG   655

Qy  2420  yLeuAlaSerGlyAspArgProProSerValSerSerValHisSerGluGluValAspCysAs  2440
          :::||||||||||||||||||||||  ::||||||||||  :: :||||||
Db   656  CTTGGCATCTGGGGACCGGACACCTTCGTTTCTCCTATGCACTCGGAG--GAGACTGC   712

Qy  2440  nArgArgThrProLeuThrAsnArgValTrpGluArgSerSer  2456
          |||||||||||||||::: :::|||||||:: :
Db   713  ACGCGCTAGGCCGTACAACACAACCCGTTTGGCAGACACACAGGCCCTATCC   761
```

RESULT 90
AW701437

```
Qy    2248  AspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGln  2267
                 |||||||  ||| |||  |||||||||||||| |||||||||||||
Db     181  GACGGGGAACAGGGGGAGCCC---AGGATGGCTTCTAAGTCCCAGGCAACACCAGCCAG   237

Qy    2268  ProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValValSerLysLys  2287
                 |||||||||||||||||||| ||||||||||||||||||||||||||
Db     238  CCGCCAGCCTTCTTCAGTAAGCTGACTGAGAGCAACTCCGCCATGGTGAAGGTGAAGAAG   297

Qy    2288  GlnGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGlyTyrAsnIleSer  2307
                 |||||||||||||| |||  |||||||||||||||||||||||||||
Db     298  CAGGAGATCAACAAGAAACTCAACACCCACAACCGGAACGAGCCAGATACAATATTGGC   357

Qy    2308  GlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyLeuMetThrTyr       2327
                 |||||||||||||||||||||||||||| |||||||||||
Db     358  CAGCCTGGGACGGAAATCTTCAACATGCCCGCCATCACTGGACCAGGCCTTATGACCTGT   417

Qy    2328  ArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArg  2347
                 |||||||| ||||||||||||||||||||||||||||  |||||| || ||
Db     418  AGAAGCCAGGCCGTGCAAGAACACGCCAGCACCAACATGGGGCTAGAGGCCATTATTAGA  477

Qy    2348  LysAlaLeuMetGlyLysTyrArgAspGlnTrpGluSerProProLeuSerAlaAsnAla  2367
                 |||||||||||||
Db     478  AAGGCACTCATG-------------------------------------------     489

Qy    2368  PheAsnProLeuAsnAlaSerAlaSerLeuProAlaAlaMetProIleThrAlaAlaAsp  2387
                 --------------------------------------------------     489
Db           ----

Qy    2388  GlyArgSerAspHisThrLeuThrSerProGlyGlyGlyLysLysAlaLysValSerGly  2407
                                              |||||||||||||||||||||
Db     490  ------------------------------GGTGGAGGTGGGAAAGCCAATGTCTCTGGC  519

Qy    2408  ArgProSerSerArgLysAlaLysSerProGlyLeuAlaSerGlyAlaSerArgGlyPro  2427
                 |||||||||||||||||||||||  |||||||||||||||||||||
Db     520  AGACCTAGCACCAGCCGAAAAGCCAAGTCGCCAGGCCTTAGCTCCGGAGCCTCGAGACCC  579

Qy,   2428  ProSerValSerSerValHisSerGluGlyAspCysAsnArgArgThrProLeuThrAsn  2447
                 |||||||||||||| |||||||||||  ||||||||||||||||||
Db     580  CCTTCTGTCTCACTCAGTACACTCCAGAAGGGGACTGCAATCGCCGAACACCACTCACCAAC  639

Qy    2448  ArgValTrpGluAsp  2452
                 |||||||||||||||
Db     640  CGTGTGTGGGAAGAC  654
```

RESULT 91
CA317171

LOCUS   CA317171   725 bp   mRNA   linear   EST 09-JUL-2003
DEFINITION UI-M-FW0-cbm-p-15-0-UI:r1 NIH_BMAP_FW0 Mus musculus cDNA clone
            IMAGE:6811744 5', mRNA sequence.
ACCESSION   CA317171
VERSION   CA317171.1   GI:24535295
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 725)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.

FEATURES   Location/Qualifiers
     source   1..725
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6811744"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FW0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
                     Site 2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is AGGGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System', supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."

ORIGIN

Alignment Scores:
Pred. No.:   1.38e-33   Length: 725
Score: 1032.00   Matches: 209
Percent Similarity: 88.52%   Conservative: 7
Best Local Similarity: 85.66%   Mismatches: 23
Query Match: 7.81%   Indels: 7
DB: 14   Gaps: 3

US-09-522-753-5 (1-2517) x CA317171 (1-725)

```
Qy    1921  LeuGlyGlyThrLeuAspGlyValTyrProThrLeuMetGluProValLeuProLys  1940
                 |  ||||||||  ||||  |||||||||||||||||||||||| |||
Db       7  TTGGTGGGCACCCTGAAGGGGTCTACCCTACCCTATGGAGCCCGTCCTGTTACCCCAAG   66

Qy    1941  GluAlaProArgValAlaArgProGluArgProArgAlaAspThrArgAlaAspThrProAlaLys  1960
                 |||  |||||||| ||||||||||||||||||||||
Db      67  GAGACCTTCGGGTCGCCCGGCCGGAGGCGCCCGTGTGGACAGCTGGCCATGCCTTCCTC  126

Qy    1961  AlaLysProProAlaArgSerGlyLeuGluProValLeuProSerLysGlySerGlu  1980
                                     ||||||||||||||||||||||
Db     127  ACCAAACCCCCGGCCCG--------GAGCCCGCCTCCTCCACCCAGCAAGAGAGTCCGAG  177

Qy    1981  ProArgProLeuValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLys  2000
                 |||||||||||||||||  |||||||||||||||||||||| ||||||
Db     178  CCCCGATCCTCAGCACCCCCAGCTCCAGCACACAGCCATCGCCCGACCCCCAGCAAAG  237

Qy    2001  AsnLeuAlaProHisHisHisAlaSerProAspProAspProAlaProProAlaSerAlaSerAsp  2020
                 ||||||||  |||||||||||||||||||||
Db     238  AACCTTGCACCCCACCATGCCAGTCCGACCCCGGCCCCGGCG---CCCACCTCGGGCCTCAGAT  294

Qy    2021  ProHisArgGluLysThrGlnSerLysProPheSerIleGluGlyLeuGluLeuArgSer  2040
                                     |||  ||||||||  |
Db     295  CTGCACCGAGAAAAGACTCAAAGTAAACCCTTTTCCATCCCAGGAATTGGAACTCCGTTCT  354

Qy    2041  LeuGlyTyrHis---GlySerSerTyrSerProGluGlyValGluProValSerProVal  2059
                 |||||  |  |||  |||||||  ||||||| |||||||||
Db     355  CTGGGTTACCCACAGTGGAGCTGGCTACAGCCCCATGGGGGTGGAGCCCATCAGCCCGGTG  414

Qy    2060  SerSerProSerLeuThrHisAspLysProLysProLysHisLeuGluLeuGluAspLys  2079
                 ||||||||||||||||||||||||||||  |||||||||||  |  |
Db     415  AGCTCCCCCAGCCTGACCCACGACAAGGGGCTCTCCAAACCTCTGGAAGAGCTAGAGAAG  474

Qy    2080  SerHisLeuGluGluLeuArgProLysGlnProGlyProValLeuGlyLeuGlyGlyGlu  2099
                 ||||  ||||||||||||||||  ||  ||  |||||  ||  |||  |  |
Db     475  AGCCACTTGGAAGGGGAGCCTGCGGGCACAAGCACCAGCAGCCCCATGAGCTCAGCGCGGAG  534
```

library, the abundant cDNAs were selected and amplified by
PCR reaction using vector region primer including T7
promotor as 5' primer and NdT14 as 3' primer. The PCR
products were used as template for synthesis of
biotinylated single stranded RNA by in vitro transcription
reaction. The synthesized RNA probes were hybridized with
antisense single stranded cDNAs prepared from original
liberary and incubated with avidin-gel. After removing
DNA-RNA hybrids by centrifuge, the subtracted cDNA
libraries were constructed by transformaion of the
remaining DNA into competent cells E. coli Top10F' with
electroporation method."

ORIGIN

Alignment Scores:
Pred. No.:　　　　1.39e-33　　　Length:　　　　614
Score:　　　　　　1030.00　　　Matches:　　　203
Percent Similarity:　72.50%　Conservative:　0
Best Local Similarity:　72.50%　Mismatches:　1
Query Match:　　　7.79%　Indels:　76
DB:　　　　　　　12　　　　Gaps:　2

US-09-522-753-5 (1-2517) x BM771666 (1-614)

Qy 2151 HisProGlnLeuSerAlaProLeuProAlaProLeuTyrSerPheProGlyAlaSer 2170
Db 3 CACCCCACAGAGCTCAGCGACCACCCCTGCCCCCCCCTCACTCCTTCCCTGGGGCAGC 62

Qy 2171 CysProValLeuAspLeuArgArgProSerAspLeuTyrLeuProProProAspHis 2190
Db 63 TGCCCCGTCTGGACCTCCGGCGCCCACCCAGTGACCTCTACCTCCGCCCCCGGACCAT 122

Qy 2191 GlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSerProGluProAsn 2210
Db 123 GGTGCCCCGGGCCCCGGCTCCCCCCACAGCAGCGAAGGGGCAAGAGTCTCCAGAGCCAAAC 182

Qy 2211 LysThrSerValLeuGlyGlyGluAspGlyIleGluProValSerProGluGlyGluGly 2230
Db 183 AAGACGTCGGTCTTGGGTGGCGGTGAGGACGTATTGAACCTGTGTCCCCACCGGAGGGC 242

Qy 2231 MetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyrArgAspHisGlyGlu 2250
Db 243 ATGACGGAGCCAGGGCACTCCCGGAGTGCTGTGTACCCGCTGCTGTACCGGGATGGGGAA 302

Qy 2251 GlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSerGlnProProAla 2270
Db 303 CAGACGGAGCCC---AGGATGGGCTCCAAGTCTCCAGGCAACACCAGCCAGCC 359

Qy 2271 PhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLysSerGlnGluIle 2290
Db 360 TTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAGAGCCAAGAGATC 419

Qy 2291 AsnLysLysLeuAsnThrHisAsnArgAsnArgAsnIleSerGlnProGly 2310
Db 420 AACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAATACAATATCGCCCAGCCTGGG 479

Qy 2311 ThrGluIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThrTyrArgSerGln 2330
Db 480 ACGGAGATCTTCAATATGCCCGCCATCACCGGAACA--------- 515

Qy 2331 AlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIleArgLysAlaLeu 2350
Db 515 --------- 515

Qy 2351 MetGlyLysTyrAspGlnTrpGluSerProProLeuSerAlaAsnAlaPheAsnPro 2370
Db 515 --------- 515

Qy 2371 LeuAsnAlaSerLeuProAlaAlaMetProIleThrAlaAlaAspGlyArgGlyArgSer 2390
Db 515 --------- 515

Qy 2391 AspHisThrLeuThrSerProGlyGlyGlyGlyLysAlaLysValSerGlyArgProSer 2410

Qy 2100 AlaAlaHisLeuProHisLeuArgProHisLeuProGluSerGlnProSerSerProLeu 2119
Db 535 GCTGCCCATCTCCCACATCTGCGGCCACTGCCCACTGCCCGGAGCCN-CCCTCATCTAGCCCACTC 593

Qy 2120 LeuGlnThrAlaProGlyHisValLysHisGlnValValThrAlaArgValValThrLeuAlaGlnHisIle 2139
Db 594 CTCCAGACTGCCCAGGCATCAAAGGTCACCAGAGGGTGGTCACCTTGGCTCAGCACATC 653

Qy 2140 SerGluValIleThrGlnAsnAspTyrThrArgHisHisProGlnGlnLeuSerAlaProLeu 2159
Db 654 AGTGAGGGTCATTACGCAGGACTACAGCGGCCACCACCCGCAGCAGCTCAGTGG--CCCCTT 712

Qy 2160 ProAlaProLeu 2163
Db 713 CCCGCCCCTCTC 724

RESULT 92
BM771666
LOCUS　BM771666　614 bp　mRNA　linear　EST 04-MAR-2002
DEFINITION　K-EST0055566 S7SNU719s1 Homo sapiens cDNA clone S7SNU719s1-21-D11
5', mRNA sequence.
ACCESSION　BM771666
VERSION　BM771666.1　GI:19101281
KEYWORDS　EST.
SOURCE　Homo sapiens (human)
ORGANISM　Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　1 (bases 1 to 614)
AUTHORS　Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE　21C Frontier Korean EST Project 2001
JOURNAL　Unpublished (2002)
COMMENT　Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 21 row: D column: 11
High quality sequence stop: 614.
FEATURES　Location/Qualifiers
source　1..614
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S7SNU719s1-21-D11"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/cell_line="SNU-719"
/lab_host="Top10F'"
/clone_lib="S7SNU719s1"
/note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library. After analyzing and
sequencing about 2,000 ~ 3,000 colonies in original cDNA

```
Db    516 ------------------GTGGCGGCGGGGAAGGCCAAGGTCTGGCAGACCCAGC 554
              |||||||||||||||||
Qy   2411 SerArgLysAlaLysSerProAlaGlyLeuAlaSerGlyAspArgProProSerVal 2430
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db    555 AGCCGAAAGCCAAGTCCCCGGCCGGCCTGGCATCTGGGGACCGGCCACCCTGTC 614

RESULT 93
BF606925
LOCUS       MY2_000117 Mouse 9-day fetus cDNA library MPMGp559    810 bp    mRNA    linear    EST 01-APR-2001
DEFINITION  cDNA clone MPMGp559C0355 5', mRNA sequence.
ACCESSION   BF606925
VERSION     BF606925.1  GI:13503489
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 810)
AUTHORS     Yahyawi,M., Hennig,S., Neidhardt,L., Radelof,U., Hermann,B.G.,
            Lehrach,H. and O'Brien,J.
TITLE       Detection of a high number of novel genes in a 9-day mouse embryo
            cDNA library normalised by oligonucleotide fingerprinting
JOURNAL     Unpublished (2001)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            Randomly selected clones.
            PCR Primers
            FORWARD: 5'-GAGCTATTCCAGAAGTAGTGA-3'
            BACKWARD: 5'-TAATACGACTCACTATAGGG-3'
            Seq primer: 5'-ATTTAGGTGACACTATAG-3'
            High quality sequence stop: 810.
FEATURES             Location/Qualifiers
     source          1..810
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="MPMGp559C0355"
                     /tissue_type="whole embryo"
                     /dev_stage="embryonic 9-day"
                     /lab_host="E.coli DH10B MaxEfficiency (Gibco cells)"
                     /clone_lib="Mouse 9-day fetus cDNA library MPMGp559"
                     /note="vector= PSVSport1; Site 1: NotI; Site 2: SalI;
                     Library preparation by oligo dT priming of RNA. Clones can
                     be ordered from the Resource Center in Berlin,
                     http://www.rzpd.de."
ORIGIN

Alignment Scores:
Pred. No.:                         1..810
Score:               2.4e-33       Length:               810
Percent Similarity:  1027.50       Matches:              215
Best Local Similarity: 81.00%      Conservative:          11
Query Match:         77.06%        Mismatches:            36
DB:                  7.78%         Indels:                19
                     10            Gaps:                   6

US-09-522-753-5 (1-2517) x BF606925 (1-810)

Qy   2088 ProLysGlnProGly--ProValLysLeuGlyLeuGlyGluAlaAlaHisLeuProHisLeu 2106
              ||||:||:|||||||||      ::|
Db     18 CCGGAATTCCCGGGTCGACCCACGCGT-----------------CCGCCCACG 53
Qy   2107 ArgProLeuProGluSerGlnProLeuSerSerProLeuLeuGlnThrAlaProGlyVal 2126
              ||||||                  ||||||||||||||||||||||||||:::
Db     54 CGTCCG-------------CCCTCATCTAGCCCACTCCTCCAGACTGCCCCAGGCATC 98
Qy   2127 LysGlyHisArgValValThrLeuAlaGlnHisHisSerGluValIleThrGlnAsp 2146
```

```
Db     99 AAAGGTACCACGAGGGTGGTCACCTTGGCTCAGCACATCAGTGAGGTCATTACGCAGAGAC 158
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2147 TyrThrArgHisHisProGlnLeuSerAlaProLeuProLeuTyrSerPhe 2166
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 TACACGGCGCCACCACCCAGCAGCTCAGTGCGCCCCCTTCCGCGCCCCTCTACTCCTTT 218
Qy   2167 ProGlyAlaSerCysProValLeuAspLeuArgArgProSerArgProAspLeuTyrLeuPro 2186
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 CCCGGAGCCAGCTGCCCTGTCCGTTCTTCGGCGGCCCACCCAGTGCCTCCTCTACCTCCA 278
Qy   2187 ProProAspHisGlyAlaProAlaArgGlySerProHisSerGluGlyGlyLysArgSer 2206
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 CCCCCCGACCATGGCACCCCAGCCCGGGGATCCCCCACAGTGAAGGGGCAAAAGGTCC 338
Qy   2207 ProGluProAsnLysThrSerValLeuGlyLeuGlyLysGlyLysGluGlyProValSer 2226
              |||::||::||:|||||||||||
Db    339 CCAGATCCCAGCAAAACATCGGTCCTGGCAGCAGGAGGATGGCCATTGAGCCTGTGTCC 398
Qy   2227 ProProGluGlyMetThrGluProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
              ||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db    399 CCACCAGAGGGCATGACTGAGCCAGGACATGCTCGGAGCCTGCGTATCCACTGCTGTAT 458
Qy   2247 ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 CGAGACGGGGAACAGGGGCCAGCCC---AGGATGGGCTCTAAGTCTCCAGGCACCACCAGC 515
Qy   2267 GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys 2286
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    516 CAGCCGCCAGCCTTCTTCAGTAAGCTGACTGAGAGCAACTCCGCCATGGTGAAGTCGAAC 575
Qy   2287 LysGlnGluIleAsnLysLysLeuAsnThrArgAsnArgAsnGluProGluTyrAsnIle 2306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    576 AAGCAGGAGATCAACATGAAACTCATCACCCCACATCCGTACGAGCCCAGAATACAATATT 635
Qy   2307 SerGlnProGlyThrGlyIlePheAsnMetProAlaIleThrGlyThrGlyLeuMetThr 2326
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    636 GGCCAGCCTGGGACGGAAATCTTCAACATGCCCGCCATCACTGGACCAGGCCTTATGTA- 694
Qy   2327 TyrArgSerGlnAlaValGlnGluHisAlaSerThrAsnMetGlyLeuGluAlaIleIle 2346
              |||||||||         ::          ::|||||||||:::::||
Db    695 CTGTATAAGAGCCAGGCGGTGCATGAC---ACGCAGCACATCATCTGGCTAGAGGCATTATTA 751
Qy   2347 ArgLysAlaLeuMetGlyLysTyrAspGlnTrpGluSerProLeuSerSerAla 2365
              :::
Db    752 TAAAGGACC---ATGGGTAAATATGATCACTGCCAAGAGCC-CCGCCGCTCGGCGCC 804

RESULT 94
AI523558/c
LOCUS       AI523558        612 bp    mRNA    linear    EST 13-APR-1999
DEFINITION  th08c09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2117680 3'
            similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
            ;contains MER22.b2 MSR1 repetitive element ;, mRNA sequence.
ACCESSION   AI523558
VERSION     AI523558.1  GI:4437693
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 612)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
```

```
FEATURES          found through the I.M.A.G.E. Consortium/LLNL at:
   source         www-bio.llnl.gov/bbrp/image/image.html
                  Insert Length: 857   Std Error: 0.00
                  Seq primer: -40UP from Gibco
                  High quality sequence stop: 488.
                  1..612
                  Location/Qualifiers
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:2117680"
                  /tissue_type="B-cell, chronic lymphotic leukemia"
                  /lab_host="DH10B"
                  /clone_lib="NCI_CGAP_CLL1"
                  /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                  polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                  was primed with a Not I - oligo7dr) primer (5'
                  TGTTACCAATCTGAAGTGGAGCGGCCGCATTGCTTTTTTTTTTTTTTTTTTTTTT
                  T 3'); double-stranded cDNA was ligated to Eco RI
                  adaptors (Pharmacia), digested with Not I and cloned into
                  the Not I and Eco RI sites of the modified pT7T3 vector.
                  Library is normalized, and was constructed by Bento
                  Soares and M.Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.:              2.23e-33        Length:         612
Score:                  1025.00         Matches:        198
Percent Similarity:     97.06%          Conservative:   0
Best Local Similarity:  97.06%          Mismatches:     6
Query Match:            7.76%           Indels:         1
DB:                                     Gaps:           0

US-09-522-753-5 (1-2517) x AI523558 (1-612)

Qy   1905  SerProValArgProAlaAlaThrPhePheProAlaThrHisCysProLeuGlyGlyThr 1924
           ||||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||
Db    612  TCACCCGGTTCGCCGGGTGCCACCATTCCCACN-GCCACCACTGGCCGGCACC        554

Qy   1925  LeuAspGlyValTyrProThrLeuMetGluProValLeuProLysGluAlaProArg 1944
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553  CTCGATGGGGTCACCCTACCCCATGGAGCCCGTCTTGCTCCCCAAGGAGGCCCCGTG  494

Qy   1945  ValAlaArgProGluArgProArgAlaAspThrGlyHisAlaPheLeuAlaLysProPro 1964
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493  GTCGCCCGGCCAGAGCGCGCCCGAGCAGACACCGGCCATGCCTTCCTCCGCCAAGCCCCCA 434

Qy   1965  AlaAlaArgSerGlyLeuGluProAlaSerProSerLysGlySerGluProArgProLeu 1984
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433  GCCCGCTCCGGGCTGGAGCCCGCCTCTCCCCCAGCAAGGGCTCGGAGCCCCGGCCCCTA 374

Qy   1985  ValProProValSerGlyHisAlaThrIleAlaArgThrProAlaLysAsnLeuAlaPro 2004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373  GTGCCTCCTGTCTCGGCCACGCCACCATCGCCCGCACCCNTGCGAAGAACCTCGCACCT 314

Qy   2005  HisHisAlaSerProAlaProProAlaProAlaSerAlaSerAspProHisArgArgGlu 2024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313  CACCACGCCAGCCCGGACCCGGCCCCGGCCTCCGCNTCCGACCCGCACCGGGGA      254

Qy   2025  LysThrGlnSerLysProPheSerIleGlnGluLeuGluLeuArgSerLeuGlyTyrHis 2044
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253  AAGACTCAAAGTAAACCCTTTTCCATCCAGGAACTGAACTCCGTTCTCGGGTTACCAC 194

Qy   2045  GlySerSerTyrSerProGlyGlyValGluProValSerProValSerProSerLeu 2064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193  GGCAGCAGCTACAGCCCCGGAAGGGGTGAGCCCCGTGAGCCCTGTGAGCTCACCCAGTCTG 134

Qy   2065  ThrHisAspLysGlyLeuHisGluGluAspLysSerHisGluGlyGly 2084
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    133  ACCCACGACAAGGGGCTCCCCAGCACCTGGACAAGAGCCACCTGGAGGGG 74

Qy   2085  GluLeuArgProLysGlnProGlyProValLysLeuGlyAlaAlaAlaHisLeuPro 2104
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     73  GAGCTGCGGCCCAAGCAGCCAGGTCCCGTGAAGCTTGGCGGGGAGGCCGGCCCACCTCCCA 14
```

```
Db     73  GAGCTGCGGCCCAAGCAGCCAGGTCCCGTGAAGCTTGGCGGGGAGGCCGGCCCACCTCCCA 14
           |||||||||||
Qy   2105  HisLeuArgPro 2108
           |||||||||||
Db     13  CACCTGCGGCCG 2
```

```
RESULT 95
BF339827
LOCUS       BF339827         602038834F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4186688
                            898 bp    mRNA    linear    EST 22-NOV-2000
DEFINITION  602038834F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4186688
            5', mRNA sequence.
ACCESSION   BF339827
VERSION     BF339827.1    GI:11286286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D.
            cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9508 row: c  column: 09
            High quality sequence stop: 765.

FEATURES          Location/Qualifiers
   source         1..898
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4186688"
                  /tissue_type="glioblastoma with EGFR amplification"
                  /lab_host="DH10B (T1 phage-resistant)"
                  /clone_lib="NCI_CGAP_Brn64"
                  /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                  Average insert size 1.57 kb. Constructed by Life
                  Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Alignment Scores:
Pred. No.:                              Length:         898
Score:                  5.23e-33        Matches:        214
Percent Similarity:     1020.50         Conservative:   33
Best Local Similarity:  82.06%          Mismatches:     41
Query Match:            71.10%          Indels:         15
DB:                     7.72%           Gaps:           5
                        10

US-09-522-753-5 (1-2517) x BF339827 (1-898)

Qy    194  LysLysLysGlnGlnGlnGluGluGluAlaAlaAlaLysProProGluProGlyLysPro 213
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3  AAAAAGAAACAACAACAGCTTGAAGAAGAGGCAGCTAAACCTCCTGAGCCTGAGGAGCCC 62

Qy    214  ValSerProProIleGluSerLysHisArgSerLeuValGlnIleIleTyrAspGlu 233
           |||||||| ||||||| ::|||   ||||||  |||||||||| ||||::|||
Db     63  GTGTCCCCTCCTCCTGTGGAGCAGAAAACACCGAGTATTGTCCAAATTATTATGATGAG 122

Qy    234  AsnArgLysAlaGluAlaGluAlaHisArgArgIleLeuGlyLeuProLysGlnValGlu 253
           |||||||| ||||||||||||||||||||  ||||::|| ||| |:::||||
Db    123  AATCGGAAAAAAGCAGAAGAAGCTCATAAATTTTGAAGGTCTTGGCCAAAAGTTGAA 182

Qy    254  LeuProLeuTyrAsnGlnProSerAspThrArgGlnTyrHisGluAsnIleLysIleAsn 273
           |||||||||||||||  :|||||||  |||:::||||::||||
Db    183  CTGCCACTGTATACCAGCCATCAGATACCAGATACCAAGGTGTACCATGAGAACATCAAGACAAAC 242
```

```
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: John Ngai. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seg primer: T7
High quality sequence stop: 497.
```

FEATURES   Location/Qualifiers
    source   1..637
                   /organism="Danio rerio"
                   /mol_type="mRNA"
                   /db_xref="taxon:7955"
                   /clone="IMAGE:4967317"
                   /sex="mixed male and female"
                   /tissue_type="brain"
                   /dev_stage="adult"
                   /lab_host="E. coli DH10B"
                   /clone_lib="zebrafish adult brain"
                   /note="Vector: pZIPLOX; Site 1: NotI; Site 2: SalI;
                   Original library was constructed in lambdaZIPLOX. Mass
                   excision of the cDNA library was performed to yield
                   pZIPLOX plasmids. Insert check was done in original
                   library."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 637
BI428444
Score: 4.77e-33   Matches: 195
Percent Similarity: 96.24%   Conservative: 10
Best Local Similarity: 91.55%   Mismatches: 7
Query Match: Indels: 1
DB: 12   Gaps: 1

US-09-522-753-5 (1-2517) x BI428444 (1-637)

```
Qy  260 ProSerAspThrArgGlnTyrHisGluAsnIleLysIleLysAsnGlnAlaMetArgLysLys 279
            CCATCTGATACCAAGCAGTACCATGAGAACATCAAAATAAATCAGGCAATGAGGAAAAAG
Db    2 CCATCTGATACCAAGCAGTACCATGAGAACATCAAAATAAATCAGGCAATGAGGAAAAAG 61

Qy  280 LeuIleLeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysGlnLysPheCys 299
            CTCTTTCTCTATTTCAAAAGAAGAAATCATGCTCGCAAACAGTGGAACAGAAGTTCTGC
Db   62 CTCTTTCTCTATTTCAAAAGAAGAAATCATGCTCGCAAACAGTGGAACAGAAGTTCTGC 121

Qy  300 GlnArgTyrAspGlnLeuMetGluAlaLeuGluLeuLysLysValGluLysIleGluAsnAsn 319
            CAGCGCTATGACCAGCTAATGGAGGCCTTGGAGAGAAGTTGAGCGATTCGAGAATAAC
Db  122 CAGCGCTATGACCAGCTAATGGAGGCCTTGGAGAGAAGTTGAGCGATTCGAGAATAAC 181

Qy  320 ProArgArgAlaLysSerLysGluSerLysValArgGluTyrTyrArgLysGlnPheProGlu 339
            CCACGACGCAGAGCCAAAGAGAGTAAAGTGCGGGAGTACTACAGAAAAGCAGTTTCCTGAG
Db  182 CCACGACGCAGAGCCAAAGAGAGTAAAGTGCGGGAGTACTACAGAAAAGCAGTTTCCTGAG 241

Qy  340 IleArgLysGlnArgGluLeuGlnLeuGlnArgMetGlnSerArgArgValGlyArgArgGlySer 359
            ATCCGCAAACAGAGAGAGCTGCAGGAGCGCATGCAGAGTAGAGTGGCTCAGCGTGGTGGG
Db  242 ATCCGCAAACAGAGAGAGCTGCAGGAGCGCATGCAGAGTAGAGTGGCTCAGCGTGGTGGG 301

Qy  360 GlyLeuSerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeu 379
            GGTCTGAGC---TCCGCAGCCCGCAGTGAACATGAGGTTTCTGAGATCATCGATGGCATC
Db  302 GGTCTGAGC---TCCGCAGCCCGCAGTGAACATGAGGTTTCTGAGATCATCGATGGCATC 358

Qy  380 SerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAlaValIleProProMetLeu 399
            TCAGAGCAGGAGAACCTCGAGAAGCAGATGCGTCAGTTAGCTGTGATTCCTCCCATGCTG
Db  359 TCAGAGCAGGAGAACCTCGAGAAGCAGATGCGTCAGTTAGCTGTGATTCCTCCCATGCTG 418

Qy  400 TyrAspAlaAspGlnArgIleLysPheIleAsnGlyLeuMetAlaAspPro 419
            TTTGATGCCGAACAGCAGAGATAAAGTTCATCAACATGAAGTGGCTGATGGATGATCCT
Db  419 TTTGATGCCGAACAGCAGAGATAAAGTTCATCAACATGAAGTGGCTGATGGATGATCCT 478

Qy  420 MetLysValTyrLysAspArgArgGlnValMetAsnMetTrpSerGluGlnGluLysGluThr 439
```

(left column)

```
Qy  274 GlnAlaMetArgLysLeuIleLeuTyrPhe-LysArgArgAsnHisAlaAlaArgLysGl 293
Db  243 CAGGCTGATGAGGAAAAACTCATTTTATTTTTTACAAAGAAGAAATCATGCAGAGAAACA 302

Qy  293 nTrpLysGlnLysPheCysGlnArgTyrAspGlnArgMetGluAlaLeuGluLysLysVa 313
Db  303 AAGGGAACAAAAAATCTGCCAGCGTATGATCAGCTCATGGAGGCATGGGAGAAAAAAGT 362

Qy  313 lGluArgIleGluAsnAsnProArgArgAlaLysSerLysGluSerLysValArgGluTyrTy 333
Db  363 GGACAGAATAGAAAATAATCCTCGGAGAAAGCTAAAGAAAGCAAAACAAGGGAATACTA 422

Qy  333 rGluLysGlnPheProGluIleArgLysGlnArgGluLeuGlnLeuGlnArgMetGlnSerAr 353
Db  423 TGAAAAGCAGTTTCCAGAAATTCGAAAACAAAGACAGCAGAGAAAGATTTCAG---CG 479

Qy  353 gValGlyGlnArgGlySerGlyLeuSerMetSerAlaAlaArgSerGluHisGluValSe 373
Db  480 AGTTGGGCAGAGGGGAGCTGGTCTTCAGCCACCAT-GCTAGGAGTGAGCATGAGATTTC 538

Qy  373 rGluIleIleAspGlyLeuSerGluGlnGluAsnLeuGluLysGlnMetArgGlnLeuAl 393
Db  539 TGAAATTATTGATGGGCTCTCTGAGCAGGAGGAATAATGAGAAACAAATGCGGCAGCTCTC 598

Qy  393 aValIleProProMetLeuTyrAspAlaAspGlnArgIleLysPheIleAsnGlyLeuMetAs 413
Db  599 TGTGATTCCACCTATGATGTTTGATGCAGAACAAGAGCAGTCCAGTTCATTAACATGAA 658

Qy  413 nGlyLeuMetAlaAspProMetLysValTyrLys-AspArgArgGlnValMetAsnMetTrpS 433
Db  659 TGGGCTTATGGAGGACCCTATGAAAGTGTTTAACAGATAGGCAGTTTATGAATGT-TGGA 717

Qy  433 erGluGlnGluLysGluThrPheArgGluArgLysPheMetGlnHisProLysAsnPheGlyL 453
Db  718 CTGACCAGGAACAGGACAATCTTCAGGACAAGTAT---CAGGCATCCAAAAAACTTGGAC 774

Qy  453 euIleAlaSerPheLeuGluArgLysThrValAlaGluCysValLeu---TyrTyrTyrL 472
Db  775 TATGCTCACCTGG----------------GAAGGAGAGTGTCCGATGGGTTCGTTTACAA 819

Qy  472 euThrLysLysAsnGluAsnTyrLysSerLeuValArgArgSerTyrArgArgArgGly 491
Db  820 TTTAACCAGAAAATGAG--------ATTTAAGCTCGTCGAGGAATATGGAACGCGGGGG 869
```

RESULT 96
BI428444
LOCUS   BI428444   637 bp   mRNA   linear   EST 26-JUL-2002
DEFINITION fr84a07.yl zebrafish adult brain Danio rerio cDNA clone
            IMAGE:4967317 5' similar to TR:Q9WU43 Q9WU43 SILENCING MEDIATOR OF
            RETINOIC ACID AND THYROID HORMONE RECEPTOR BETA. [1] ;, mRNA
            sequence.
ACCESSION   BI428444
VERSION   BI428444.1   GI:15205676
KEYWORDS   EST.
SOURCE   Danio rerio (zebrafish)
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 637)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee.R., Li,F., Marra,M.,
            Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE   WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
  COMMENT   Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800

```
Db         |||||||||:::
         4 GACACAAAGTCTACCAGCAGAAAACATCAAAACAACCAGGTGATGAGGAGAAAAAGCTAATA    63

Qy   282 LeuTyrPheLysArgArgAsnHisAlaArgLysGlnTrpLysPheCysGlnArg    301
         |||:::||||||||||||||||||||   ::||||||||
Db    64 CTATTTTTTAAAAGAGAATCATGCAGAAATCATGCAGAAAAATCTGCCAACGT    123

Qy   302 TyrAspGlnLeuMetGluAlaLeuGluLysLysValGluArgIleGluAlaAsnProArg    321
         ||||||||:||:||||||||||||||||::|:::|||
Db   124 TACGATCAGCTGATCGAAGCATGGGAAAAAAAGTTGACAGGATAGAAAATAATCCACGC    183

Qy   322 ArgArgAlaLysGluSerLysValArgGlnValTyrTyrGluLysGlnPheProGluIleProArg    341
         |||||||:||:|:::||||||||
Db   184 AGGAAAGCTAAGGAGCAAAACAAGGGAATACTATGAGAAACAATTTCCAGAAATTCCGG    243

Qy   342 LysGlnArgGluLeuGlnLeuArgMetGlnSerArgValGlyValArgGlySerGlyLeu    361
         |||||||||              ||::||||||
Db   244 AAGCAAAGAGAACAGCAAGAGCGATTTCAG---AGAGTTGGTCAAGAGAGGTGCTGGACTT    300

Qy   362 SerMetSerAlaAlaArgSerGluHisGluValSerGluIleIleAspGlyLeuSerGlu    381
         |||   ::   ||||||:||:::
Db   301 TCAGCAACTATTGCTAGAAGCGAACATGAGATTTCTGAAATCATTGATGGGCTTTCTGAG    360

Qy   382 GlnGluAsnLeuGluLysGlnMetArgGlnLeuLeuAlaValIleProProMetLeuTyrAsp    401
         |||||||                ||::::||||||:||||||||:||||
Db   361 CAGGAGAATAATGAAAAACAAATGCGTCAGCTCTCAGTCATTCCTCCCATGAATGTTTGAT    420

Qy   402 AlaAspGlnAsnArgIleLysPheIleAsnMetAsnGlyLeuMetAlaAspProMetLys    421
         |||||:::||:::||||||::||||||||||||
Db   421 GCAGACAGAGACCAGTGAAGTTTATTAATAATGATGGTCTGATGGAGGATCCCATGAAA    480

Qy   422 ValTyrLysAspArgArgGlnValMetAsnMetTrpSerGluGlnLysGluThrPheArg    441
         ||:::||||||||
Db   481 GTTTTATAAAGACAGACAATTCATGAATGTCTGGACCGACCACGAGAAGAGATTTTTAA    540

Qy   442 GluLysPheMetGlnAsnHisProLysAsnPheGlyLeuIleAlaSerPheLeuGluArgGlyLys    461
         ||||||||::||||||||||||||||||||
Db   541 GAAAAGTTTGTCCAACATCCCAAGAACNTNGGTCTAATTGCTTCCTACCTGGAACGAAAG    600

Qy   462 ThrValAlaGluCysValLeuTyrTyrTyrLeuThrLysLysAsnTyrLysSer    481
         ||| ::|||||||||||||||||::||||||||||::||||
Db   601 AATGTTCCTGACTGTGTACTCTATTATTATTTGACTAAGAAAAATGAAAACTATAAAGCC    660

Qy   482 LeuValArgArgSerTyr---ArgArgArgGlyLysSerGlnGlnGlnGlnGln    500
         |||:::::||||||||||||||||||||
Db   661 CTTGTGCGAGGAGAATTATGTAAACGCAGAGGGAGAAACCAGCAACAA--------    708

Qy   501 GlnGlnGlnGlnGlnGlnProMetProArgSerGlnGluGlyLys    520
         ||||||||||||||||||||
Db   709 -----------------------------ATGGCTCGTCCTTCTCAAGAGAGAAAA    735

Qy   521 AspGluLysGlu    524
         |||:::||
Db   736 GTAGAAAGAAAA    747
```

RESULT 98
AW674918
LOCUS AW674918 612 bp mRNA linear EST 11-APR-2000
DEFINITION bb30f04.y1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:2984767 5'
  similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
  ;, mRNA sequence.
ACCESSION AW674918
VERSION AW674918.1 GI:7540153
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 612)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: bb30f04.x1

```
Db         ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::|||
       479 ATGAAGGTCTACAAAGACAGACAAGGTCATGATGCGGGATCATGTGACGAACAGGAGGACACC    538

Qy   440 PheArgGluLysPheMetGlnHisProLysAsnPheGlyLeuAlaSerPheLeuGlu    459
         |||||:||::|||             ||||||||||
Db   539 TTCCGTGAGAAGTTTATCCAGCACCACCCAAAAACTTTGCCCGATAGCCTCTTTTCTGGAG    598

RESULT 97
BU143568
LOCUS  BU143568  760 bp  mRNA  linear  EST 25-NOV-2002
DEFINITION 603229009F1 CSEQCHL26 Gallus gallus cDNA clone ChEST221b4 5', mRNA
  sequence.
ACCESSION  BU143568
VERSION  BU143568.1  GI:25360169
KEYWORDS  EST.
SOURCE  Gallus gallus (chicken)
ORGANISM  Gallus gallus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
  Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 760)
AUTHORS  Boardman,P.E., Sanz-Ezguerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
  Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE  A Comprehensive Collection of Chicken cDNAs
JOURNAL  Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE  22335534
PUBMED  12445392
COMMENT  Contact: Simon Hubbard
  Department of Biomolecular Sciences
  University of Manchester Institute of Science and Technology
  (UMIST)
  PO Box 88, Manchester, M60 1QD, UK
  Tel: 01612008930
  Fax: 01612360409
  Email: Simon.Hubbard@umist.ac.uk.
FEATURES  Location/Qualifiers
source  1..760
  /organism="gallus gallus"
  /mol_type="mRNA"
  /strain="Compton Line 15I"
  /db_xref="taxon:9031"
  /clone="ChEST221b4"
  /sex="Female"
  /dev_stage="adult"
  /lab_host="DH10B"
  /clone_lib="CSEQCHL26"
  /note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
  EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)
  (Stratagene) vector to accommodate cDNA produced with the
  T-trimmed protocol (Construction of uni-directionally
  cloned cDNA libraries from messenger RNA for improved 3'
  end DNA sequencing by Glenn Fu, et al. U.S. Patent #
  6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI.
  Ligate in double stranded adaptor containing BsgI and
  BamHI sites [5'ggccggtgagcccggatccgaaaaaag]
  [5'aattcctttttcggatccgggctgcacgc]"
ORIGIN

Alignment Scores:
Pred. No.:    Length:  760
Score:  5.8e-33  Matches:  197
Percent Similarity: 1017.50 Conservative: 28
Best Local Similarity: 85.23% Mismatches: 22
Query Match:  74.62% Indels:  17
DB:  7.70%  Gaps:  3
  13

US-09-522-753-5 (1-2517) x BU143568 (1-760)

Qy   262 AspThrArgGlnTyrHisGluAsnIleLysIleIleAsnGlnAlaMetArgLysLysLeuIle    281

Qy   460 ArgLysThrValAlaGluCysValLeuTyrTyrTyrLeu    472
         |||||:||||||||||||||
Db   599 AGAAAGACCGGTGGCTGAGTGTGTCCTTCTACTACCTG    637
```

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40RP from Gibco
High quality sequence stop: 408.

FEATURES
source   1..612
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:2984767"
                /cell_line="MGC36"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_10"
                /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.5 kb. Library prepared by Life
                Technologies."

ORIGIN

Alignment Scores:
Pred. No.:   6.38e-33   Length: 612
Score: 1014.00   Matches: 190
Percent Similarity: 95.10%   Conservative: 4
Best Local Similarity: 93.14%   Mismatches: 10
Query Match: 7.67%   Indels: 0
DB: Gaps: 0

US-09-522-753-5 (1-2517) x AW674918 (1-612)

Qy 2144 ThrGlnAspTyrThrArgHisHisProGlnLeuSerAlaProLeuProLeu 2163
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ACACAGGACTACACCCGGCACCACCCCACAGCAGCTCAGCGACGCCACCCTGCCCGCCCCCCTC 60

Qy 2164 TyrSerPheProGlyAlaSerCysProValLeuAspLeuArgArgProSerAspLeu 2183
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TACTCCTTCCCTGGGGCCAGCTGCCCCGTCCTGGACCTCCGCCGCCCCAGCTGACCTC 120

Qy 2184 TyrLeuProProAspHisGlyAlaArgGlySerProHisSerGluGlyGly 2203
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TACCTCCCGCCCGACCATGTGCCCCGGCCCGTGCTCCCCCACAGCGAAGGGGGC 180

Qy 2204 LysArgSerProGluProAsnLysThrSerValLeuGlyGlyGlyIleGlu 2223
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGAGGTCTCCAGAGCCAAACAAGACGTCGGTTCTTGGGTGGTGGTGATTGAA 240

Qy 2224 ProValSerProGluGlyMetThrGlyProGlyHisSerArgSerAlaValTyrPro 2243
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCTGTGTCCCCAGCGAGGGGCATGACGGGACCAGGGCACTCCCGGAGTGCTGTGTACCCG 300

Qy 2244 LeuLeuTyrArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGly 2263
         |||||||||||||||||||||||||||||:|||||||||||||||||||
Db   301 CTGCTGTACCGGGATGGGGAACAGACCGGAGCCCAGCAAGATGGGCTCCAAGTCTCCAGGC 360

Qy 2264 AsnThrSerGlnProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetVal 2283
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AACACCAGCCAGCCCGCCTTCTTCAGCAAGCTGACCGAGAGCAACTCCGCCATGGTC 420

Qy 2284 LysSerLysLysGlnProGluIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGlu 2303
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AAGTCCAAAAAGCAAGAGATCAACAAGAAGCTGAACACCCACAACCGGAATGAGCCTGAA 480

Qy 2304 TyrAsnIleSerGlnProGlyThrGluIlePheAsnMetProAlaIleThrGlyThrGly 2323
         ||||||||||||||||||||||||||||::::|||:::|||||||
Db   481 TACAATATCAGCCAGCCTGGGACAGAGATCTTCAATATGCCAGCCATGACCAGAACATGC 540

Qy 2324 LeuMetThrTyrArgSerGlnAlaValGlnGluHisHisAlaSerThrAsnMetGlyLeuGlu 2343

Db   541 CTTATGACCTATAGAAGCAGGCAGGCAGAAGCAGCAGAGGCAGAACATCCCTGCATCACATAGGGCTGGTA 600
         ||||||||||||| |||||||| |||||||| |::::|||||||

Qy 2344 AlaIleAlaIleArg 2347
         ||||||||||
Db   601 GGCATAATTAGA 612

RESULT 99
BE675456
BE675456

LOCUS   7f09d11.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3294165 3'
           610 bp   mRNA   linear   EST 08-SEP-2000
DEFINITION similar to TR:O00613 O00613 T3 RECEPTOR-ASSOCIATING COFACTOR-1. [1]
           ;, mRNA sequence.
ACCESSION BE675456
VERSION   BE675456.1 GI:10035997
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 610)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
           M.D., Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL, send email to:
           info@image.llnl.gov
           Seq primer: -40UP from Gibco
           High quality sequence stop: 431.
FEATURES
source   1..610
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3294165"
                /tissue_type="B-cell, chronic lymphotic leukemia"
                /lab_host="DH10B"
                /clone_lib="NCI_CGAP_CLL1"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site 1: Eco RI; Site 2: Eco RI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5'
                TGTTACCAATCTGAGTGGAGCGGCGCCATTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
                T 3']; double-stranded cDNA was ligated to Eco RI
                adaptors (Pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of the modified pT7T3 vector.
                Library is normalized, and was constructed by Bento
                Soares and M.Fatima Bonaldo."

ORIGIN

Alignment Scores:
Pred. No.:   7.69e-33   Length: 610
Score: 1012.00   Matches: 192
Percent Similarity: 96.04%   Conservative: 2
Best Local Similarity: 95.05%   Mismatches: 8
Query Match: 7.66%   Indels: 0
DB: Gaps: 0

US-09-522-753-5 (1-2517) x BE675456 (1-610)

Qy 2107 ArgProLeuProGluSerGlnProSerSerProLeuGlnThrAlaProGlyVal 2126
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 CGGCCGCTGCCTGAGAGCCAGCCCTCTCCAGCCCGCTGCTCCAGACCGCCCCCAGGGGTC 61

Qy 2127 LysGlyHisGlnArgValValThrLeuAlaGlnHisIleSerGluValIleThrGlnAsp 2146
         |||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    62 AAAGGTCACCAGCGGGTGGTCACCCTGCCCAGCACATCAGTGAGGTCATCACACAGGAC 121

Qy  2147 TyrThrArgHisProGlnIleuSerAlaProLeuProAlaProLeuTyrSerPhe 2166
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TACACCCGGCACCACCCCACCAGCAGCTCAGCGACCCTGCCCCGCCCCCTCTACTCCTTT 181

Qy  2167 ProGlyAlaSerCysProValLeuAspLeuArgArgProSerAspLeuProTyrLeuPro 2186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 CCTGGGGCCAGCTGCCCCGTCCTGGACCTTCCGCCGCCCACCCAGTGACCTCTACCTCCCG 241

Qy  2187 ProProAspHisGlyAlaProAlaAlaArgGlySerProHisSerGluGlyGlyLysArgSer 2206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 CCCCCGGACCATGGTGCCCCGGCCCGTGGCCTCCCCCACAGCGAAGGGGCAAGAGGTCT 301

Qy  2207 ProGluProAsnLysThrSerValLeuGlyGlyGluAspGlyIleGluProValSer 2226
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 CCAGAGCCAAACAAGACGTCGGTCTTGGGTGGTGGTGCAGGACACGGTATTGAACCTGTGTCC 361

Qy  2227 ProProGluGlyMetThrGlyProGlyHisSerArgSerAlaValTyrProLeuLeuTyr 2246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 CCACCGGAGGGCATGACCGGACCCAGGGCACTCCCGGAGTGCTGTGTACCCGTGCTGTAC 421

Qy  2247 ArgAspGlyGluGlnThrGluProSerArgMetGlySerLysSerProGlyAsnThrSer 2266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 CGGGATGGGGAACAGACGGAGCCCATCAGGATGGGCTCCAAGTCTCCAGGCAACACCAGC 481

Qy  2267 GlnProProAlaPhePheSerLysLeuThrGluSerAsnSerAlaMetValLysSerLys 2286
              |||||||||||||||||||:::||||||||||||||||||||||||||||
Db   482 CAGCCTGCAGCCTTCTTCAGCGAGCTGACCGAGAGCAACTCCGCCATGGTCAAGTCCAAG 541

Qy  2287 LysGlnIleIleAsnLysLysLeuAsnThrHisAsnArgAsnGluProGluTyrAsnIle 2306
              |||||||| :::: |||||   ||| ||||||||||| ||||||||
Db   542 AAGCAAGAGATCATCAAGAAGCCTGACAATCACCAAACCGAATGAGCCCTGATACAATATC 601

Qy  2307 SerGln 2308
              ||||||
Db   602 AGCCAA 607
```

RESULT 100

BY763864
LOCUS　　　　BY763864　　　　　　　　　676 bp　　mRNA　　　linear　　EST 17-DEC-2002
DEFINITION　BY763864 RIKEN full-length enriched, NOD-derived CD11c +ve
　　　　　　　dendritic cells Mus musculus cDNA clone F630016H22 3', mRNA
　　　　　　　sequence.
ACCESSION　　BY763864
VERSION　　　BY763864.1　GI:27200956
KEYWORDS　　 EST.
SOURCE　　　 Mus musculus (house mouse)
ORGANISM　　 Mus musculus
　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE　　1　(bases 1 to 676)
AUTHORS　　　Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
　　　　　　　Nikaido,I., Osato,N., Saito,N., Suzuki,H., Yamanaka,I.,
　　　　　　　Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
　　　　　　　Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
　　　　　　　Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
　　　　　　　Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
　　　　　　　Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
　　　　　　　Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
　　　　　　　Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
　　　　　　　Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
　　　　　　　Kawai,H., Kawasaki,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
　　　　　　　Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
　　　　　　　Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
　　　　　　　Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
　　　　　　　Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
　　　　　　　Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
　　　　　　　Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
　　　　　　　Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
　　　　　　　Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
　　　　　　　Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
　　　　　　　Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,

Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE　　　　Analysis of the mouse transcriptome based on functional annotation
　　　　　　　of 60,770 full-length cDNAs
JOURNAL　　　Nature 420, 563-573 (2002)
MEDLINE　　　2235483
PUBMED　　　　12466851
COMMENT　　　Contact: Yoshihide Hayashizaki
　　　　　　　Laboratory for Genome Exploration Research Group, RIKEN Genomic
　　　　　　　Sciences Center(GSC), Yokohama Institute
　　　　　　　The Institute of Physical and Chemical Research (RIKEN)
　　　　　　　1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
　　　　　　　Tel: 81-45-503-9222
　　　　　　　Fax: 81-45-503-9216
　　　　　　　Email: genome-res@gsc.riken.go.jp,
　　　　　　　URL:http://genome.gsc.riken.go.jp/
　　　　　　　Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
　　　　　　　Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
　　　　　　　Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Murata,J., Kojima,Y.,
　　　　　　　Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
　　　　　　　Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
　　　　　　　Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
　　　　　　　Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
　　　　　　　Direct Submission
　　　　　　　Computational Analysis of Full-Length Mouse cDNAs Compared with
　　　　　　　Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
　　　　　　　Normalization and subtraction of cap-trapper-selected cDNAs to
　　　　　　　prepare full-length cDNA libraries for rapid discovery of new
　　　　　　　genes. Genome Res. 10 (10), 1617-1630 (2000)
　　　　　　　RIKEN integrated sequence analysis (RISA) system--384-format
　　　　　　　sequencing pipeline with 384 multicapillary sequencer. Genome Res.
　　　　　　　10 (11), 1757-1771 (2000)
　　　　　　　Computer-based methods for the mouse full-length cDNA
　　　　　　　encyclopedia: real-time sequence clustering for construction of a
　　　　　　　nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
　　　　　　　cDNA library was prepared and sequenced in Mouse Genome
　　　　　　　Encyclopedia Project of Genome Exploration Research Group in Riken
　　　　　　　Genomic Sciences Center and Genome Science Laboratory in RIKEN.
　　　　　　　Division of Experimental Animal Research in Riken contributed to
　　　　　　　prepare mouse tissues.
　　　　　　　Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
　　　　　　　Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
　　　　　　　Trust/MRC building Addenbrookes Hospital Cambridge) whose
　　　　　　　assistance we gratefully acknowledge.
　　　　　　　Please visit our web site (http://genome.gsc.riken.go.jp) for
　　　　　　　further details.
FEATURES　　　　　Location/Qualifiers
　　　source　　　1..676
　　　　　　　　　/organism="Mus musculus"
　　　　　　　　　/mol_type="mRNA"
　　　　　　　　　/strain="NOD"
　　　　　　　　　/db_xref="taxon:10090"
　　　　　　　　　/clone="F630016H22"
　　　　　　　　　/cell_type="NOD-derived CD11c +ve dendritic cells"
　　　　　　　　　/clone_lib="RIKEN full-length enriched, NOD-derived CD11c
　　　　　　　　　+ve dendritic cells"
ORIGIN

Alignment Scores:
Pred. No.:   9.03e-33   Length: 676
Score: 1011.50   Matches: 191
Percent Similarity: 90.62%   Conservative: 12
Best Local Similarity: 85.27%   Mismatches: 18
Query Match: 7.65%   Indels: 3
DB: 13   Gaps: 1

US-09-522-753-5 (1-2517) x BY763864 (1-676)

Qy 1387 ProProProSerArgLeuThrGluAlaTyrLysThrGln--------AlaLeuGly 1403
```

Mon Apr 19 08:15:09 2004

us-09-522-753-5.rst

```
Db          |||       |||        ||||||||||||||||||||||      ||||||      :::              |||||||
        4   CCCACACCACCTNAGGACCTGACTGAGACNTACAAGCNCCGGCCCCTGGACCCCTCTGGGT   63

Qy  1404    ProLeuLysLysProAlaHisGluGlyLeuValLysGluAlaThrVaLLysGluAlaGlyArg   1423
            |||||||||||||||||||      |||:::|||||||||||      :::||||||||||||
Db    64    CCCCTGAAGCTGAAGCCGACTCACGAGGGTGTGGTAGCAACTGTGAAGAAGGCGGGCCGC      123

Qy  1424    SerIleHisGluIleProArgGluLeuArgHisThrProGluLeuProLeuAlaPro   1443
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124    TCTATCCATGAGATCCCGAGAGAGGAGCTGCGCCGCCACACCTGAGCTACCCCCTGGCACCA    183

Qy  1444    ArgProLeuLysGluGlySerIleThrGlnGlyThrProLeuLysTyrAspThrGlyAla   1463
            |||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db   184    CGGCCCTCTGAAGGAGGGTTCCATCACCCAGGGCACCCCACTCAAGTACGACTCTGGGGCA    243

Qy  1464    SerThrThrGlySerLysHisAspValArgSerLeuIleGlySerProGlyArgThr   1483
            :::||||||      |||||||||||||:::||||||||||||||||||||||||||||||
Db   244    CCCTCCACTGGCACCAAGAAACACGACGTGCGCTCAATCATCGGCACCCCCGGCCGGCCT    303

Qy  1484    PheProProValHisProLeuAspValMetAlaAspAlaArgAlaLeuGluArgAlaCys   1503
            :::||||||      :::||||||||||||||||||||||||||||||||||||||||||
Db   304    TTCCCTGCCGTGCACCCCCTGGACATAATGGCTGACGCCCGGGCACTGGAGCCGTGCCTGC    363

Qy  1504    TyrGluGluSerLeuLysSerArgProGlyThrAlaSerSerGlyGlySerIleAla   1523
            |||||||||||      |||:::|||||||||||||||||||||||||||||||||||||
Db   364    TATGAAGAGAGTCTGAAGAGCCGGTCAGGGACCAGCAGTGGTGCAGGGGGCTCCATCACA    423

Qy  1524    ArgGlyAlaProValIleValProGluLeuGlyLysProArgGlnSerProLeuThrTyr   1543
            ||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424    CGTGGGGCTCCAGTCGTCGTGCCTGAACTGGGCAAGCCACGGCAAAGCCCACTGACTTAC    483

Qy  1544    GluAspHisGlyAlaProPheAlaGlyHisLeuProArgGlySerProValThrMetArg   1563
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484    GAAGACCACGGGGCACCCTTCGCTGGTCACCTGCCACGTGGCCTCCCCCGTGACCACGAGG    543

Qy  1564    GluProThrProArgLeuGlnGlyGlySerSerSerLeuSerSerLysAlaSerGlnAspArg   1583
            ||||||||||||||||||      |||||||||||||||||||||||||||||||||||||
Db   544    GAGCCCACGCCACGCCTTCAGGAAGGCAGCAGCTCCTATCCAGCAAGGCGTCCCAGGACCGG    603

Qy  1584    LysLeuThrSerThrProArgGluIleAlaLysSerProHisSerThrValProGluHis   1603
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   604    AAGCTGACATCTACACCCCGGGAGATCGCCAAGTCCCCACACAGCACTGTGCCCGAGCAC    663

Qy  1604    HisProHisPro   1607
            ||||||||||||
Db   664    CACCCTCACCCC    675
```

Search completed: April 16, 2004, 08:14:39
Job time : 12777 secs